



UNIVERSIDAD  
**PABLO<sup>D</sup>  
OLAVIDE**  
S E V I L L A

**TESIS DOCTORAL**

# Análisis del genoma de *Sphingopyxis granuli* estirpe TFA y caracterización de un nuevo elemento regulador de los genes de degradación de tetralina

**INMACULADA GARCÍA ROMERO**







TESIS DOCTORAL

**Análisis del genoma de *Sphingopyxis granuli* estirpe TFA y caracterización de un nuevo elemento regulador de los genes de degradación de tetralina**

Memoria presentada por **Inmaculada García Romero** para optar al grado de Doctora

Sevilla, noviembre de 2017

**El director de la Tesis**

**La directora de la Tesis**

**Dr. Eduardo Santero Santurino**

Catedrático

Dpto. Biología Molecular e Ingeniería Bioquímica  
Universidad Pablo de Olavide

**Dra. Belén Floriano Pardal**

Profesora Titular

Dpto. Biología Molecular e Ingeniería Bioquímica  
Universidad Pablo de Olavide



Departamento de Biología Molecular e Ingeniería Bioquímica

Programa de Doctorado en Biotecnología, Ingeniería y Tecnología Química (RD: 99/2011)



## FINANCIACIÓN Y AYUDAS PARA LA REALIZACIÓN DE ESTA TESIS DOCTORAL

Inicialmente quiero agradecer a las instituciones y fondos de financiación que han hecho posible la realización de esta Tesis doctoral, que se ha llevado a cabo en el Centro Andaluz de Biología del Desarrollo (CABD-CSIC-UPO), dentro del departamento de Biología Molecular e Ingeniería Bioquímica de la Universidad Pablo de Olavide de Sevilla. La Tesis doctoral ha sido financiada por una beca **FPI** del Ministerio de Economía y Competitividad en la convocatoria de 2012, con referencia **BES-2012-053751** y asignada al proyecto **BIO2011-24003 (*Integración de circuitos reguladores en operones de regulación compleja en bacterias de interés medioambiental*)**, cuyo Investigador Principal es Dr. Eduardo Santero Santurino.

Además se ha realizado una estancia breve en el Centro de Investigaciones Biológicas de Madrid (CIB-CSIC) financiada por una ayuda para la realización de estancias breves asociada a la beca FPI, en la convocatoria de 2014 y con referencia **EEBB-I-15-09452**.

Finalmente esta tesis también ha sido financiada mediante un proyecto del Ministerio de Economía y Competitividad en la convocatoria de 2014 y con referencia **BIO2014-57545-R (*Modelos de regulación global y específica en bacterias degradadoras de contaminantes ambientales*)**, cuyo Investigador Principal es Dr. Eduardo Santero Santurino.



## AGRADECIMIENTOS

Parecía que no iba a llegar este momento y aquí estoy, escribiendo los agradecimientos de mi Tesis y, aunque los haya dejado para el final de la escritura, no es ni mucho menos porque sean poco importantes sino más bien todo lo contrario.

Quiero empezar agradeciendo a mis directores de Tesis Eduardo y Belén por haber confiado en mí para la realización de este trabajo, por guiarme y orientarme cuando me encontraba más perdida y por todo lo que he aprendido gracias a ellos, desde que entré al laboratorio como alumna interna hasta hoy. De corazón, muchísimas gracias.

Continúo con Antonio, porque sin él una gran parte de esta Tesis no habría sido posible. Infinitas gracias por todo lo que me has enseñado sobre bioinformática, por haber estado siempre dispuesto a ayudarme y orientarme, por la paciencia cuando no eran ni uno, ni dos, ni tres...los correos que te mandaba a la semana para resolver problemas con el ensamblaje del genoma, la anotación, etc. Gracias porque contagias la pasión por tu trabajo.

De director de máster paso a directora, Paqui, muchas gracias por estar siempre dispuesta a discutir de ciencia conmigo, por escuchar mis agobios, por esas risas incontrolables y contagiosas que muchas veces he necesitado y por todos los consejos que me has dado durante todo este tiempo, gracias porque haces que se trabaje “con más ahínco si cabe”. Y continúo en el despachito, muchas gracias a Eva y Carlos por las veces que me he “girado” para preguntarles cualquier dudilla y estar siempre dispuestos a ayudar. No puedo dejar de agradecer al resto de profes: Inés, Amando, Fernan, Aroa...porque también han estado ahí para responder dudas, aconsejar, etc. En especial a Aroa le agradezco que me haya ayudado con la portada de la Tesis, ¡Es una auténtica diseñadora!

Y volviendo la vista unos años atrás, en los comienzos, en la etapa donde hay que aprender a manejarse en el laboratorio, fueron fundamentales las “niñas de micro”: Bea, Laura Terrón, Ana, Laura Ledesma, Cristina, Sofía, Ali, Elena y Helena. Muchas gracias a todas ellas por ayudarme y enseñarme en la parte experimental de mi Tesis, por los consejos y “truquillos” que solo ellas sabían para hacer los experimentos, por enseñarme a usar los equipos del laboratorio, etc. Muchos de mis experimentos llevan una parte de vosotras, millones de gracias. Y entre ellas tengo que destacar a las tetralinas. Laura, con esos videos tan divertidos que me enseñaba mientras estaba ahí en el despachito escribiendo la Tesis y yo, con mis cosas de bioinformática, despejaba un poquito la mente. Helena, ¡Ay señor la infinidad de cosas que te pregunté y te sigo preguntando casi hasta ahora! Te debo muchísimo porque me has enseñado muchas cosas y ayudado con los protocolos cuando más perdida me encontraba, además, fue utilísimo para mi Tesis colaborar contigo y que me enseñaras a construir mutantes en TFA, de corazón, te estaré eternamente agradecida. Por último Elena, que ella cree que no le hago caso en los consejos que me da pero yo le prometo que siempre los tengo en mente, muchas gracias por todos los consejos que me has dado, no solo para el trabajo en el laboratorio, sino de la vida en general y, ya más al final de mi periodo en el laboratorio, por los ratillos de café junto con Rubén.

Y como pieza fundamental del trabajo en el laboratorio quiero destacar a Guada. Para mí ha sido imprescindible, desde encontrar reactivos que se habían esfumado (pero luego ella iba y los veía, en modo madre) hasta ayudarme con miles de protocolos (gracias por tu paciencia



conmigo para extraer genómico de TFA en aquellos momentos de desesperación). Esta Tesis, por supuesto, lleva gran parte de tus enseñanzas y por ello te estoy muy agradecida.

Y es hora de pasar a los más pequeños: Toni, Yolanda, Rubén, Yuzu, Rocío, Eli...con todos ellos he pasado algún ratillo divertido en el laboratorio, en las jornadas doctorales, etc. En especial, mis compis de desayuno, Rubén y Yuzu, han sido fundamentales para continuar el día con más energía, por contarnos nuestras cosas de fuera y dentro del lab, reírnos, porque estuvieron ahí para ver cómo una ventisca casi me lleva volando por la terraza del CABD y un sinfín de anécdotas divertidas. Rubén, en ti voy a detenerme, porque te has convertido en un gran amigo y ha sido maravilloso coincidir contigo aquí. Espero que nuestra amistad perdure muuuchos años más. Además, como ya te he dicho mil veces, eres un miniIP, me has aconsejado mucho con los experimentos y eres una fuente de sabiduría, no tengo duda de que tu carrera científica será excelente.

Durante mi estancia en el Centro de Investigaciones Biológicas (CIB) de Madrid quiero mostrar mi agradecimiento a Eduardo Díaz y Juan Nogales por permitir que desarrollase parte del trabajo de esta Tesis con ellos. A Juan le debo todo lo que sé sobre modelos metabólicos y que me haya ilusionado tantas veces durante el tiempo que estuve allí con todas las posibilidades de estudio que ofrece la construcción y el manejo de modelos a escala genómica. Muchas gracias a todos mis compañeros de allí que me hicieron pasar una etapa muy amena en esta andadura. Y siguiendo con centros de investigación externos también quiero agradecer a mi directora Belén Floriano por su estancia investigadora en el laboratorio de Cynthia Sharma y Jörg Vogel en la Universidad de Würzburg (Alemania), porque gran parte de su trabajo allí también está reflejado en esta Tesis.

Y fuera del laboratorio quiero agradecer a mis preciosas Isa y Ro, por animarme tanto, transmitirme tanta positividad y por los momentos juntas que siempre son revitalizantes. A mis compis del crucero, Estefi, Patri, Vero, Irene y Fran con los que he compartido muchos momentos felices y han estado ahí para escuchar mis preocupaciones e inquietudes relacionadas con la Tesis. Al resto de Biotech: Noelia, Sara G, Sara N, Pablis, Ana Belén, Sandra...con los que he compartido momentos agradables y charlado sobre las experiencias de cada uno en nuestra etapa laboral post-licenciatura.

A mis amig@s de Berrocal, por escucharme con tanto interés cuando me preguntaban que cómo me iba con la Tesis, en el segundo uno se arrepentían y me decían: ¡Ruby no me entero de ná!, pero a mí me alegraba y daba ánimos contarles mi trabajo. Muchas gracias, amig@s.

Y por último, con mi familia la palabra agradecimiento se queda corta. A mi hermano, gracias por esos purés tan ricos cuando llegaba al piso a las tantas sin ganas de cocinar y por cuidarme tan bien los años que hemos vivido juntos en Sevilla. A mis padres, a ellos les debo todo lo que soy, han confiado en mí, me han animado y apoyado desde el minuto uno en el que pisé Sevilla para empezar la carrera hasta el final de mi Tesis. Os quiero mucho y os agradezco todo. Y a Juan, mi familia también, llegaste a mi vida al inicio de esta aventura y si no nos ha separado la Tesis no nos separará nada, porque mira que he canalizado en ti todo el estrés acumulado en ocasiones y tú has sabido tranquilizarme y hacerme reflexionar para que no lo viese todo tan dramático. Muchas gracias porque junto a ti todo es más fácil.

¡MUCHAS GRACIAS A TODOS!

*“A mis padres, mi hermano, mi abuelo Juan y en memoria de  
mis abuelos Javier, Santana y Pilar”*



## **ÍNDICE**





## INTRODUCCIÓN

<b>1. Los compuestos aromáticos y su biodegradación</b>	3
<b>2. La familia <i>Sphingomonadaceae</i></b>	5
<b>3. Degradación de tetralina en <i>S. granuli</i> TFA</b>	7
3.1. La ruta de degradación de tetralina	8
3.2. Regulación específica de los genes <i>thn</i> de <i>S. granuli</i> TFA	10
3.3. Represión catabólica de los genes <i>thn</i> de <i>S. granuli</i> TFA	13
<b>4. Secuenciación, ensamblaje y anotación de los genomas bacterianos</b>	14
4.1. Tecnologías para la secuenciación de genomas	16
4.2. Ensamblado de las secuencias procedentes de las tecnologías de secuenciación masiva	16
4.3. Anotación y caracterización de la información genómica	17
4.3.1. Anotación estructural de genes codificantes de proteínas	18
4.3.2. Anotación funcional de genes codificantes de proteínas	19
4.3.3. Detección y anotación funcional de ARNs no codificantes	22
<b>5. Pequeños ARNs reguladores no codificantes en bacterias</b>	24
5.1. Tipos de ARNs reguladores y mecanismos de acción	24
5.1.1. ARNs reguladores codificados en <i>cis</i>	25
5.1.2. ARNs reguladores codificados en <i>trans</i>	26
5.2. Procesos biológicos regulados por pequeños ARNs en bacterias	28
<b>6. Biología de sistemas</b>	29
6.1. Genómica comparada	29
6.2. Reconstrucción a escala genómica del metabolismo microbiano	31
6.2.1. Construcción y análisis de modelos metabólicos	32
<b>OBJETIVOS</b>	37

## MATERIALES Y MÉTODOS

<b>1. Estirpes, plásmidos, oligonucleótidos y condiciones de cultivo</b>	43
1.1. Estirpes y plásmidos	43
1.2. Oligonucleótidos	45
1.3. Medios y condiciones de cultivo	46
1.4. Conservación del material biológico	48
<b>2. Secuenciación del genoma de <i>S. granuli</i> TFA</b>	48
2.1. Tecnología de secuenciación 454	48

2.2.	Ensamblaje de las lecturas y unión de los fragmentos generados .....	49
<b>3.</b>	<b>Herramientas bioinformáticas y bases de datos para la anotación del genoma de <i>S. granuli</i> TFA y la reconstrucción metabólica .....</b>	<b>51</b>
3.1.	Anotación del genoma de <i>S. granuli</i> TFA.....	51
3.1.1.	Anotación estructural y funcional de genes codificantes.....	52
3.1.2.	Anotación de ARNs no codificantes .....	53
3.1.3.	Identificación del origen de replicación .....	53
3.1.4.	Identificación de islas genómicas, profagos y secuencias CRISPR.....	54
3.2.	Reconstrucción metabólica de <i>S. granuli</i> TFA .....	54
3.2.1.	Obtención del borrador inicial del modelo .....	54
3.2.2.	Curación manual del modelo inicial.....	55
3.2.3.	Conversión del modelo a formato computacional .....	58
3.2.4.	Evaluación inicial y relleno de huecos ( <i>gap filling</i> ) en las rutas metabólicas del modelo .....	58
3.2.5.	Denominación del modelo .....	59
3.2.6.	Evaluación de la capacidad predictiva del modelo.....	59
<b>4.</b>	<b>Herramientas bioinformáticas para los análisis de genómica comparada .....</b>	<b>61</b>
4.1.	Construcción de árboles filogenéticos .....	61
4.1.1.	Concatenación de genes conservados .....	61
4.1.2.	Cálculo de la media en identidad de aminoácidos (AAI, del inglés <i>Average Amino Acid Identity</i> ) .....	61
4.1.3.	Cálculo de la media en identidad de nucleótidos (ANI, del inglés <i>Average Nucleotide Identity</i> ).....	62
4.2.	Comparación de secuencias genómicas y contexto génico .....	62
4.3.	Construcción de una matriz de proteomas .....	62
4.4.	Obtención del pangenoma y el genoma núcleo del género <i>Sphingopyxis</i> .....	63
<b>5.</b>	<b>Transferencia de ADN a estirpes bacterianas .....</b>	<b>63</b>
5.1.	Transformación.....	63
5.1.1.	Transformación de <i>E. coli</i> mediante choque térmico .....	63
5.1.2.	Electrotransformación de <i>S. granuli</i> TFA .....	64
5.2.	Conjugación triparental.....	65
<b>6.</b>	<b>Manipulación de ácidos nucleicos .....</b>	<b>66</b>
6.1.	Manipulación de ADN .....	66
6.1.1.	Aislamiento de ADN plasmídico de <i>Escherichia coli</i> .....	66
6.1.2.	Aislamiento de ADN total de <i>S. granuli</i> .....	66
6.1.3.	Electroforesis de ADN en geles de agarosa .....	66

6.1.4.	Purificación de fragmentos de ADN .....	67
6.1.5.	Clonación de fragmentos de ADN.....	67
6.1.6.	Detección de secuencias de ADN por hibridación en membrana ( <i>Southern Blot</i> ) .....	68
6.1.7.	Detección de secuencias de ADN por hibridación en colonias .....	69
6.1.8.	Reacción en cadena de la polimerasa (PCR) .....	69
6.1.9.	Hibridación y extensión de oligonucleótidos solapantes .....	70
6.1.10.	Secuenciación de pequeños fragmentos de ADN .....	71
6.1.11.	Construcción de plásmidos y de estirpes de TFA.....	71
6.2.	Manipulación de ARN.....	73
6.2.1.	Aislamiento de ARN de <i>S. granuli</i> TFA .....	73
6.2.2.	Síntesis de transcritos de ARN mediante transcripción <i>in vitro</i> .....	74
6.2.3.	Electroforesis de ARN .....	75
6.2.4.	Detección de secuencias de ARN por hibridación en membrana ( <i>Northern Blot</i> ) .....	76
6.3.	Cuantificación de ácidos nucleicos .....	77
<b>7.</b>	<b>Detección de proteínas mediante inmunodetección (<i>Western Blot</i>) .....</b>	<b>77</b>
<b>8.</b>	<b>Análisis de la expresión génica .....</b>	<b>79</b>
8.1.	Ensayos de actividad $\beta$ -galactosidasa .....	79
8.1.1.	Obtención de células para la medida de expresión de fusiones a <i>lacZ</i> .....	79
8.1.2.	Medida de la actividad $\beta$ -galactosidasa .....	80
8.2.	Experimentos de transcriptómica (dRNA-seq) .....	81
<b>9.</b>	<b>Detección y validación de la interacción ARN-ARN .....</b>	<b>82</b>
9.1.	Detección <i>in silico</i> de interacciones ARN-ARN .....	82
9.2.	Sistema reportero basado en fusiones a GFP .....	82
9.2.1.	Construcción de vectores.....	82
9.2.2.	Ensayos de fluorescencia .....	83
9.3.	Ensayo de retardo de la movilidad en gel (EMSA, ARN-ARN) .....	84
<b>10.</b>	<b>Co-inmunoprecipitación de ARNs con Hfq .....</b>	<b>85</b>

## RESULTADOS

<b>1.</b>	<b>Análisis del genoma de <i>S. granuli</i> TFA .....</b>	<b>89</b>
1.1.	Secuenciación y ensamblado.....	89
1.2.	Características generales del genoma de <i>S. granuli</i> TFA.....	94
1.2.1.	Análisis del uso de codones.....	96

1.2.2.	Identificación del origen de replicación .....	99
1.3.	Organización del genoma de <i>S. granuli</i> TFA: evidencias de transferencia horizontal, islas genómicas y profagos .....	104
1.3.1.	Identificación de regiones del genoma de <i>S. granuli</i> TFA homólogas a otros genomas. ....	104
1.3.2.	Identificación de islas genómicas .....	106
1.3.3.	Presencia de plásmidos integrados .....	107
1.3.4.	Identificación de profagos y secuencias CRISPR .....	108
<b>2.</b>	<b>Genómica comparada</b> .....	<b>110</b>
2.1.	Adscripción filogenética de <i>S. granuli</i> TFA al género <i>Sphingopyxis</i> mediante árboles filogenéticos .....	110
2.2.	Adscripción filogenética de <i>S. granuli</i> TFA al género <i>Sphingopyxis</i> mediante una matriz de proteomas .....	113
2.3.	Determinación de la especie de <i>S. granuli</i> TFA .....	115
2.4.	Análisis del pangenoma y el genoma núcleo del género <i>Sphingopyxis</i> .....	116
<b>3.</b>	<b>Reconstrucción metabólica de <i>S. granuli</i> TFA</b> .....	<b>119</b>
3.1.	Descripción general del modelo metabólico de <i>S. granuli</i> TFA .....	122
3.2.	Validación del modelo metabólico iLG738 de <i>S. granuli</i> TFA .....	125
3.2.1.	Validación cualitativa de la utilización de fuentes de carbono .....	125
3.2.2.	Validación cuantitativa de la utilización de fuentes de carbono .....	130
3.2.3.	Validación del proceso de respiración anaerobia .....	132
<b>4.</b>	<b>Identificación de ARNs no codificantes en <i>S. granuli</i> TFA</b> .....	<b>135</b>
<b>5.</b>	<b>Caracterización del pequeño ARN regulador SuhB</b> .....	<b>136</b>
5.1.	Estudio de la conservación de la secuencia y el contexto génico de SuhB en el género <i>Sphingopyxis</i> .....	136
5.2.	Expresión diferencial de <i>suhB</i> .....	139
5.3.	Construcción de un mutante carente del pequeño ARN SuhB .....	143
5.3.1.	Caracterización del crecimiento de la estirpe MPO643 .....	145
5.3.2.	Caracterización de la inducción de los genes <i>thn</i> en el mutante MPO643 en medio mínimo con tetralina .....	147
5.3.3.	Caracterización de la inducción de los genes <i>thn</i> en el mutante MPO643 en condiciones de represión catabólica .....	148
5.3.4.	Complementación funcional de la estirpe MPO643 .....	149
5.3.5.	Comparación del fenotipo de represión catabólica de las estirpes MPO643 y MPO209-1002 .....	150

5.3.6.	Caracterización de la expresión de <i>suhB</i> en el mutante MPO209 (inserción en <i>phaC</i> ) .....	151
5.3.7.	Búsqueda de posibles dianas del pequeño ARN SuhB .....	152
5.3.8.	Estudio de la regulación de la expresión de <i>thnR</i> mediada por SuhB .....	154
5.3.9.	Validación <i>in vitro</i> de la interacción entre SuhB y la región 5'UTR de <i>thnR</i> .....	156
5.3.10.	Análisis de la implicación de Hfq en la interacción entre <i>thnR</i> y SuhB ....	158

## DISCUSIÓN

<b>1.</b>	<b>Organización y características del genoma de <i>S. granuli</i> TFA .....</b>	<b>163</b>
1.1.	Adscripción filogenética de TFA a la especie <i>S. granuli</i> , estilo de vida y nicho ecológico .....	168
<b>2.</b>	<b>Estudio de las capacidades metabólicas a través del modelo <i>ΔG738</i> .....</b>	<b>170</b>
2.1.	Degradación de alcohol de polivinilo (PVA) .....	172
2.2.	Degradación de compuestos aromáticos .....	174
2.3.	Síntesis de ectoína .....	174
2.4.	Producción de gránulos de PHB y su efecto sobre el crecimiento .....	176
<b>3.</b>	<b>Implicación de un nuevo elemento en la regulación de los genes <i>thn</i> .....</b>	<b>177</b>

<b>CONCLUSIONES.....</b>	<b>185</b>
--------------------------	------------

<b>ANEXOS.....</b>	<b>189</b>
--------------------	------------

<b>BIBLIOGRAFÍA.....</b>	<b>217</b>
--------------------------	------------





## ÍNDICE DE TABLAS

<b>Tabla 1</b>	Estirpes bacterianas utilizadas durante la realización de este trabajo	43
<b>Tabla 2</b>	Plásmidos utilizados durante la realización de este trabajo	43
<b>Tabla 3</b>	Oligonucleótidos utilizados durante la realización de este trabajo (salvo los empleados en el proceso de ensamblaje del genoma de TFA)	45
<b>Tabla 4</b>	Fuentes de carbono probadas en TFA	47
<b>Tabla 5</b>	Concentraciones de antibióticos y otros agentes químicos añadidos a los medios de cultivo	48
<b>Tabla 6</b>	Criterios para indicar el nivel de fiabilidad de la inclusión de una reacción en el modelo metabólico	56
<b>Tabla 7</b>	Composición de la reacción de biomasa de TFA	57
<b>Tabla 8</b>	Características más relevantes del ensamblado conseguido con cada uno de los ensambladores automáticos	89
<b>Tabla 9</b>	Características de la secuencia y anotación del genoma de <i>Sphingopyxis granuli</i> TFA	95
<b>Tabla 10</b>	Regla del tambaleo codón-anticodón	96
<b>Tabla 11</b>	Uso de codones en <i>Sphingopyxis granuli</i> TFA	97
<b>Tabla 12</b>	Islas genómicas cuyas coordenadas solapan en las predicciones de los dos métodos utilizados en IslandViewer 3	106
<b>Tabla 13</b>	Características de los tres profagos identificados con la herramienta PHAST	108
<b>Tabla 14</b>	Resultado del análisis de ANIm en porcentaje de identidad y, entre paréntesis, porcentaje de secuencia incluida en el alineamiento	115
<b>Tabla 15</b>	Resultado del alineamiento del gen codificante del ARNr 16S en el servidor EzBioCloud	116
<b>Tabla 16</b>	Comparación de los distintos modelos metabólicos publicados de bacterias miembros de la clase $\alpha$ -proteobacteria	124
<b>Tabla 17</b>	Lista de metabolitos testados en medio mínimo como única fuente de carbono para TFA	125
<b>Tabla 18</b>	Proteínas implicadas en la degradación de 3-fenilpropanoato y 3-(3-hidroxifenil)propanoato en TFA	128
<b>Tabla 19</b>	Parámetros para el cálculo del consumo de las distintas fuentes de carbono en TFA obtenidos a partir de dos réplicas biológicas	131
<b>Tabla 20</b>	Tasa de crecimiento ( $\text{h}^{-1}$ ) de TFA silvestre y de la estirpe MPO209 (mutante en <i>phaC</i> )	134
<b>Tabla 21</b>	ARNs no codificantes anotados en el genoma de TFA	135

## ÍNDICE DE FIGURAS

<b>Figura 1</b>	Molécula de tetralina	7
<b>Figura 2</b>	Ruta de degradación de tetralina en <i>S. granuli</i> TFA	9
<b>Figura 3</b>	Organización génica de los genes implicados en la degradación de tetralina en <i>S. granuli</i> TFA	10
<b>Figura 4</b>	Modelo de regulación específica de la expresión de los genes <i>thn</i> en presencia de la molécula inductora (A) o sustratos no metabolizables (B)	12
<b>Figura 5</b>	Expresión de la fusión <i>thnC::lacZ</i> en TFA silvestre y en los ocho candidatos de la mutagénesis con transposón	14
<b>Figura 6</b>	Número de genomas de bacterias y arqueas secuenciados cada año y subidos a NCBI	15
<b>Figura 7</b>	Representación de pequeños ARNs codificados en <i>cis</i>	25
<b>Figura 8</b>	Representación de pequeños ARNs codificados en <i>trans</i>	26
<b>Figura 9</b>	Estructura general de los pequeños ARNs reguladores codificados en <i>trans</i>	27
<b>Figura 10</b>	Representación esquemática del pan-genoma de una especie	30
<b>Figura 11</b>	Definición gráfica del concepto de modelos basados en restricciones	34
<b>Figura 12</b>	Bote diseñado para el suministro de tetralina en fase gas en cultivos líquidos de <i>Sphingopyxis granuli</i>	47
<b>Figura 13</b>	Esquema de la cascada enzimática que tiene lugar en la pirosecuenciación	49
<b>Figura 14</b>	Esquema del flujo de trabajo seguido para la obtención de las distintas hipótesis de unión entre los fragmentos del genoma de TFA	51
<b>Figura 15</b>	Esquema del proceso de anotación	51
<b>Figura 16</b>	Obtención de cultivos para medida de la actividad $\beta$ -galactosidasa de la fusión traduccional <i>thnC::lacZ</i>	79
<b>Figura 17</b>	Obtención de cultivos para medida de la actividad $\beta$ -galactosidasa de la fusión transcripcional <i>P<sub>suhB</sub>::lacZ</i>	80
<b>Figura 18</b>	Representación de las distintas combinaciones de plásmidos para el ensayo de interacción entre 5'UTR- <i>thnR</i> y SuhB	83
<b>Figura 19</b>	Acumulación de lecturas en el genoma artificial de TFA, las cuales se localizan en el fragmento ctg220-217	91

<b>Figura 20</b>	Representación esquemática de las acumulaciones de lecturas en el fragmento ctg220-217 y de las repeticiones de secuencias en los extremos de los 4 fragmentos finales del genoma de TFA	92
<b>Figura 21</b>	Resultado del <i>Southern Blot</i> empleando la sonda del extremo 5' del fragmento ctg229-248 sobre digestiones del ADN cromosómico de TFA	94
<b>Figura 22</b>	Resultado de la predicción de Ori-Flinder	100
<b>Figura 23</b>	Análisis de la región genómica que contiene el OriC	102
<b>Figura 24</b>	Representación del contenido en GC (en negro) y la desviación GC, (en verde y morado), en el genoma de TFA	103
<b>Figura 25</b>	Representación circular del genoma de TFA y comparación con los genomas de otras $\alpha$ -proteobacteria	105
<b>Figura 26</b>	Análisis filogenético de TFA. Árbol filogenético construido a partir de las secuencias concatenadas de los genes conservados 16S, <i>atpD</i> y <i>rpoB</i>	111
<b>Figura 27</b>	Análisis filogenético de TFA. Dendrograma basado en la media de identidad en aminoácidos	112
<b>Figura 28</b>	Comparación de proteomas entre miembros de las familias <i>Sphingomonadaceae</i> y <i>Erythrobacteraceae</i>	114
<b>Figura 29</b>	Representación del pangenoma y el genoma núcleo del género <i>Sphingopyxis</i>	117
<b>Figura 30</b>	Clasificación de las proteínas pertenecientes al genoma núcleo del género <i>Sphingopyxis</i>	118
<b>Figura 31</b>	Representación con diagramas de Venn de las reacciones (Rxn), metabolitos (Met) y genes (Genes) incorporados en el borrador inicial de modelo de TFA	120
<b>Figura 32</b>	Clasificación de las reacciones del modelo <i>i</i> IG738	122
<b>Figura 33</b>	Representación esquemática de las principales rutas catabólicas incluidas en el modelo <i>i</i> IG738	126
<b>Figura 34</b>	Representación esquemática de las rutas de degradación de 3-(3-hidroxi-fenil)propanoato y 3-fenilpropanoato según la información disponible en la base de datos MetaCyc	127
<b>Figura 35</b>	Representación esquemática de la posible ruta de degradación del alcohol de polivinilo (PVA) en TFA, considerando un suministro de PQQ	130
<b>Figura 36</b>	Velocidad de crecimiento de TFA en diferentes fuentes de carbono	131
<b>Figura 37</b>	Optimización de la reacción de biomasa en <i>i</i> IG738, utilizando 40 mM de $\beta$ -HB como fuente de carbono y 20 mM de nitrato ( $\text{NO}_3$ ) como último aceptor de electrones	132

<b>Figura 38</b>	Predicción del crecimiento de TFA por el modelo <i>ilG738</i>	133
<b>Figura 39</b>	Detección in silico y experimental del pequeño ARN <i>SuhB</i> en TFA	136
<b>Figura 40</b>	Estructura de la secuencia del pequeño ARN <i>SuhB</i>	136
<b>Figura 41</b>	Conservación del contexto génico de <i>suhB</i> en los genomas de bacterias del género <i>Sphingopyxis</i>	137
<b>Figura 42</b>	Alineamiento de la región intergénica que contiene a <i>suhB</i> en TFA y la región homóloga en el resto de <i>Sphingopyxis</i>	138
<b>Figura 43</b>	Conservación de la estructura de <i>SuhB</i> en las distintas <i>Sphingopyxis</i>	139
<b>Figura 44</b>	Expresión diferencial de <i>suhB</i> en $\beta$ -HB y tetralina	140
<b>Figura 45</b>	Detección de <i>SuhB</i> mediante <i>Northern Blots</i>	140
<b>Figura 46</b>	Expresión de <i>suhB</i> en células de TFA crecidas en $\beta$ -HB 40 mM o THN	141
<b>Figura 47</b>	Expresión de <i>suhB</i> en células de TFA crecidas en $\beta$ -HB 40 mM, ácido sebácico 16 mM o THN	142
<b>Figura 48</b>	Expresión de <i>suhB</i> en células de TFA	142
<b>Figura 49</b>	Localización de <i>suhB</i> en el genoma de TFA (A) y esquema de la construcción del plásmido pMPO1150 (B)	143
<b>Figura 50</b>	<i>Southern Blot</i> de ADN cromosómico de TFA silvestre y los candidatos $\Delta$ <i>suhB::kan</i>	144
<b>Figura 51</b>	Curvas de crecimiento de T1002 (TFA; <i>thnC::lacZ</i> ) y MPO643 ( $\Delta$ <i>suhB::kan</i> ; <i>thnC::lacZ</i> ) en distintos medios	146
<b>Figura 52</b>	Curva de crecimiento y expresión de la fusión traduccional <i>thnC::lacZ</i> en MM con tetralina en las estirpes T1002 (TFA; <i>thnC::lacZ</i> ) y MPO643 ( $\Delta$ <i>suhB::kan</i> ; <i>thnC::lacZ</i> )	147
<b>Figura 53</b>	Curva de crecimiento y expresión de la fusión traduccional <i>thnC::lacZ</i> en MM con tetralina y $\beta$ -HB 40 mM en las estirpes T1002 (TFA; <i>thnC::lacZ</i> ) y MPO643 ( $\Delta$ <i>suhB::kan</i> ; <i>thnC::lacZ</i> )	148
<b>Figura 54</b>	Complementación funcional del mutante MPO643 con el vector pMPO1152	149
<b>Figura 55</b>	Curva de crecimiento y expresión de la fusión traduccional <i>thnC::lacZ</i> en MM con tetralina y $\beta$ -HB 40 mM en las estirpes T1002 (TFA; <i>thnC::lacZ</i> ), MPO643 ( $\Delta$ <i>suhB::kan</i> ; <i>thnC::lacZ</i> ) y MPO209-1002 (mutante de inserción en <i>phaC</i> ; <i>thnC::lacZ</i> )	150
<b>Figura 56</b>	Curvas de crecimiento de las estirpes MPO641 ( <i>PsuhB::lacZ</i> TFA silvestre) y MPO644 (mutante de inserción en <i>phaC</i> ; <i>PsuhB::lacZ</i> ) en MM con $\beta$ -HB 40 mM, THN o ambas fuentes de carbono	151



<b>Figura 57</b>	Expresión de la fusión transcripcional <i>PsuhB::lacZ</i> de las estirpes MPO641 (TFA; <i>PsuhB::lacZ</i> ) y MPO644 (mutante de inserción en <i>phaC</i> ; <i>suhB::lacZ</i> ) en MM con $\beta$ -HB 40 mM, THN o ambas fuentes de carbono.	152
<b>Figura 58</b>	Representación de la interacción entre SuhB y la región 5'UTR del ARNm de <i>thnR</i> predicha por IntaRNA	154
<b>Figura 59</b>	Detección de ThnR por <i>Western Blot</i> y niveles de expresión de la fusión <i>thnC::lacZ</i> a las 10 horas de inducción	155
<b>Figura 60</b>	Interacción predicha por IntaRNA entre SuhB y las versiones silvestre y mutante de la región 5'UTR de <i>thnR</i>	156
<b>Figura 61</b>	Ensayos de retardo de movilidad en gel del transcrito de la región 5'UTR de <i>thnR</i> con SuhB	157
<b>Figura 62</b>	Co-inmunoprecipitación (CoIP) de SuhB y <i>thnR</i> con Hfq-3xFLAG	159
<b>Figura 63</b>	<i>Northern Blot</i> con la sonda complementaria a SuhB en las distintas etapas de la co-inmunoprecipitación	159
<b>Figura 64</b>	Fluorescencia de la fusión traduccional <i>thnR::sfgfp</i> normalizada con respecto al control pXG0	160
<b>Figura 65</b>	Convergencia de la degradación de distintos compuestos aromáticos en el metabolito 2-oxopent-4-enotato	174
<b>Figura 66</b>	Conservación del pequeño ARN SuhB en el género <i>Sphingopyxis</i>	177
<b>Figura 67</b>	Modelo de regulación de la expresión de los genes <i>thn</i> en condiciones de inducción (A) y represión catabólica (B)	183



## ABREVIATURAS

<b>aa</b>	aminoácido
<b>AAI</b>	media en identidad de amino ácidos ( <i><u>A</u>verage <u>A</u>mino acid <u>I</u>dentity</i> )
<b>ADN</b>	ácido desoxirribonucleico
<b>ADNc</b>	ADN codificante
<b>ANI</b>	media en identidad de nucleótidos ( <i><u>A</u>verage <u>N</u>ucleotide <u>I</u>dentity</i> )
<b>Ap</b>	ampicilina
<b>APS</b>	adenosin 5' fosfosulfato
<b>ARN</b>	ácido ribonucleico
<b>ARNm</b>	ARN mensajero
<b>ARNr</b>	ARN ribosómico
<b>ARNt</b>	ARN transferente
<b>ATP</b>	adenosín trifosfato
<b>BLAST</b>	<i><u>B</u>asic <u>L</u>ocal <u>A</u>lignment <u>S</u>earch <u>T</u>ool</i>
<b>BOF</b>	función objetivo de la biomasa ( <i><u>B</u>iomass <u>O</u>bjective <u>F</u>unction</i> )
<b>BSTFA</b>	N,O-bis(trimetilsilil)trifluoroacetamida
<b>BTEX</b>	acrónimo de <u>B</u> enceno, <u>T</u> olueno, <u>E</u> tilbenzeno y <u>X</u> ileno
<b>BWA</b>	<i><u>B</u>urrows-<u>W</u>heeler <u>A</u>ligner</i>
<b>CCD</b>	dispositivo de carga acoplada ( <i><u>C</u>harge-<u>C</u>oupled <u>D</u>evice</i> )
<b>CCR</b>	<i><u>C</u>arbon <u>C</u>atabolite <u>R</u>epression</i>
<b>CM</b>	modelo de covarianza ( <i><u>C</u>ovariance <u>M</u>odel</i> )
<b>Cm</b>	cloranfenicol
<b>cm</b>	centímetros
<b>CoA</b>	coenzima A
<b>COBRA</b>	reconstrucción y análisis basado en restricciones ( <i><u>C</u>Onstraint-<u>B</u>ased <u>R</u>econstruction and <u>A</u>alysis</i> )
<b>CRISPR</b>	<i>clustered regularly interspaced short palindromic repeats</i>
<b>CTP</b>	citidín trifosfato
<b>D.O.<sub>600</sub></b>	<u>D</u> ensidad <u>Ó</u> ptica medida a 600 nm
<b>dATP</b>	desoxiadenosín trifosfato
<b>dCTP</b>	desoxicitosín trifosfato
<b>DEPC</b>	dietilpirocarbonato
<b>dFBA</b>	FBA dinámico
<b>dGTP</b>	desoxiguanosín trifosfato
<b>DMSO</b>	dimetil sulfóxido
<b>dNTP</b>	desoxinucleótido trifosfato

<b>dRNA-seq</b>	secuenciación diferencial de ARN ( <i><u>D</u>ifferential <u>R</u>NA <u>S</u>equencing</i> )
<b>dTTP</b>	desoxitimidín trifosfato
<b>EDTA</b>	ácido etilendiaminotetraacético
<b>EMSA</b>	ensayo de reardo de movilidad en gel
<b>FAD</b>	flavín adenín dinucleótido oxidado
<b>FADH<sub>2</sub></b>	flavín adenín dinucleótido reducido
<b>FBA</b>	Análisis equilibrado de flujos ( <i><u>F</u>lux <u>B</u>alanced <u>A</u>nalysis</i> )
<b>Fd</b>	factor de dilución
<b><i>g</i></b>	aceleración de la gravedad terrestre
<b>g</b>	gramos
<b>GC-MS</b>	cromatografía de gases y espectrometría de masas
<b>GEM</b>	modelo metabólico a escala genómica ( <i><u>G</u>Enome-scale <u>M</u>odel</i> )
<b>GEMRE</b>	reconstrucción a escala genómica del metabolism ( <i><u>G</u>Enome-scale <u>M</u>etabolic <u>R</u>econstruction</i> )
<b>Gm</b>	gentamicina
<b>GPR</b>	relación gen-proteína-reacción
<b>GTP</b>	guanosín trifosfato
<b>h</b>	hora
<b>ic ó IC</b>	inversa complementaria
<b>IG</b>	isla genómica
<b>IPTG</b>	isopropil-β-D-tiogalactósido
<b>kb</b>	kilobase
<b>Km</b>	kanamicina
<b>kV</b>	kilovoltio
<b>L</b>	litros
<b>LB</b>	Medio <u>L</u> uria- <u>B</u> ertani
<b>lb</b>	límite bajo ( <i><u>l</u>ower <u>b</u>ounds</i> )
<b>M</b>	molar
<b>mA</b>	miliamperios
<b>Mb</b>	megabases
<b>mg</b>	miligramos
<b>min</b>	minutos
<b>MM</b>	<u>M</u> edio <u>M</u> ínimo
<b>mM</b>	milimolar
<b>mm</b>	milimetros
<b>MML</b>	medio rico de crecimiento para TFA

<b>mmol</b>	milimol
<b>NAD(P)</b>	nicotinamida-adenín-dinucleótido o dinucleótido (fosfato) oxidado
<b>NAD(P)H</b>	nicotinamida-adenín-dinucleótido o dinucleótido (fosfato) reducido
<b>NCBI</b>	<u>N</u> ational <u>C</u> enter for <u>B</u> iotechnology <u>I</u> nformation
<b>NEB</b>	<u>N</u> ew <u>E</u> ngland <u>B</u> iolabs
<b>ng</b>	nanogramos
<b>NGS</b>	<u>N</u> ext- <u>G</u> eneration <u>S</u> equencing
<b>nm</b>	nanómetros
<b>nM</b>	nanomolar
<b>°C</b>	grados centígrados
<b>ONPG</b>	o-nitrofenil galactopiranósido
<b>ORF</b>	Pauta abierta de lectura ( <u>O</u> pen <u>R</u> eadin <u>G</u> <u>F</u> rame)
<b>p/v</b>	relación peso/volumen
<b>PAGE</b>	electroforesis en gel de poliacrilamida ( <u>P</u> oly- <u>A</u> cryl <u>a</u> mide <u>G</u> el <u>E</u> lectrophoresis)
<b>P<sub>B</sub></b>	promotor del operón B
<b>pb</b>	pares de bases
<b>PBS</b>	tampón fosfato salino
<b>P<sub>c</sub></b>	promotor del operón C
<b>PCR</b>	reacción en cadena de la polimerasa
<b>PEG</b>	polietilenglicol
<b>P<sub>H</sub></b>	promotor interno del operón B, aguas arriba de <i>thnH</i>
<b>PHAs</b>	hidrocarburos aromáticos policíclicos
<b>PHB</b>	polihidroxibutirato
<b>P<sub>lac</sub></b>	promotor del operón <i>lac</i>
<b>P<sub>L</sub>lacO1</b>	promotor P <sub>L</sub> con el operador <i>lac</i>
<b>P<sub>L</sub>tetO1</b>	promotor P <sub>L</sub> con el operador <i>tet</i>
<b>P<sub>M</sub></b>	promotor del operón M
<b>pmoles</b>	picomoles
<b>PNK</b>	polinucleótido kinasa
<b>PPi</b>	pirofosfato
<b>PQQ</b>	pirroloquinolina quinon
<b>P<sub>R</sub></b>	promotor interno del operón C, aguas arriba de <i>thnR</i>
<b>P<sub>suhB</sub></b>	promotor del gen <i>suhB</i>
<b>P<sub>tac</sub></b>	Promotor <i>tac</i>
<b>PVA</b>	alcohol de polivinilo
<b>RBS</b>	<u>r</u> ibosomal <u>b</u> inding <u>s</u> ite

<b>rpm</b>	revoluciones por minuto
<b>Sac</b>	sacarosa
<b>SBML</b>	<i><u>S</u>ystems <u>B</u>iology <u>M</u>arkup <u>L</u>anguage</i>
<b>SD</b>	Shine-Dalgarno
<b>SDS</b>	dodecilsulfato sódico
<b>SFF</b>	<i><u>S</u>tandard <u>F</u>lowgram <u>F</u>ormat</i>
<b>SMRT</b>	<i><u>S</u>ingle <u>M</u>olecule <u>R</u>eal <u>T</u>ime</i>
<b>SOB</b>	medio <i><u>S</u>uper <u>O</u>ptimal <u>B</u>roth</i>
<b>SOC</b>	medio <i><u>S</u>uper <u>O</u>ptimal broth with <u>C</u>atabolic repressor</i>
<b>sRNA</b>	pequeño ARN regulador
<b>SSC</b>	citrato de sodio salino
<b>Str</b>	estreptomicina
<b>T7</b>	fago T7
<b>TAE</b>	tris-acetato-EDTA
<b>TB</b>	medio <i><u>T</u>errific <u>B</u>roth</i>
<b>TBE</b>	tris-borato-EDTA
<b>TBS</b>	tampón tris salino
<b>Tc</b>	tetraciclina
<b>TE</b>	tris-EDTA
<b>TEMED</b>	N,N,N',N'-tetrametiletildiamina
<b>TER</b>	tris-EDTA con ribonucleasa
<b>TEX</b>	<i><u>T</u>erminator 5' phosphate-dependent <u>e</u>xonuclease</i>
<b>tg</b>	tiempo de generación
<b>THN</b>	tetralina
<b>TIV</b>	transcripción <i>in vitro</i>
<b>TMCS</b>	Trimetilclorosilano
<b>TSS</b>	inicio de la transcripción ( <i><u>T</u>ranscription <u>S</u>tart <u>S</u>ite</i> )
<b>TTBS</b>	tampón tris salino con Tween 20
<b>U</b>	unidades
<b>ub</b>	límite alto ( <i><u>u</u>pper <u>b</u>ounds</i> )
<b>UTP</b>	uridín trifosfato
<b>UTR</b>	<i>untranslated region</i>
<b>V</b>	voltios
<b>v/v</b>	relación volumen/volumen
<b>wt</b>	estirpe silvestre ( <i><u>w</u>ild <u>t</u>ype</i> )
<b>X-gal</b>	5-bromo-4-cloro-3-indol-β-D-galactopiranósido

<b>β-HB</b>	β-hidroxibutirato
<b>μCi</b>	microcurio
<b>μL</b>	microlitros
<b>μM</b>	micromolar





## **INTRODUCCIÓN**

---



La capacidad de los microorganismos para utilizar como fuente de carbono y energía multitud de compuestos orgánicos naturales o producidos por la actividad humana, resulta indispensable para la continuidad de la vida en la tierra. Además, esta versatilidad metabólica favorece la rápida adaptación de los microorganismos a los cambios en el medio y les permite habitar en nichos hostiles que resultan letales para la vida de organismos superiores. La degradación microbiana de compuestos recalcitrantes, tanto naturales como xenobióticos, suscita un gran interés debido a que su persistencia en el medio puede constituir un grave problema medioambiental y de salud. Dentro de estos compuestos recalcitrantes se encuentran los hidrocarburos aromáticos policíclicos (PAHs, acrónimo en inglés), cuyas propiedades tóxicas y carcinogénicas hacen especialmente preocupante su acumulación en los ecosistemas. Entre los microorganismos capaces de biodegradar PAHs se han descrito numerosos representantes de la familia *Sphingomonadaceae*, a la cual pertenece la bacteria objeto de estudio en esta tesis. La secuenciación y anotación de genomas completos de bacterias con la habilidad de degradar compuestos contaminantes es el primer paso para conocer en más detalle, no ya su capacidad degradadora concreta, sino los mecanismos globales de regulación génica que controlan la expresión de los genes implicados en el catabolismo de dichos contaminantes, así como la fisiología de estos microorganismos y sus capacidades metabólicas generales, información que puede integrarse en modelos metabólicos para comprender y predecir su comportamiento en distintas condiciones ambientales.

### 1. Los compuestos aromáticos y su biodegradación

Los compuestos aromáticos son abundantes en la naturaleza, representando aproximadamente el 20 % de la biomasa de la Tierra, y estructuralmente diversos (Díaz y col., 2013). Se encuentran en todos los organismos vivos en forma de los aminoácidos fenilalanina, triptófano y tirosina y pueden llegar a constituir un cuarto de la biomasa de las plantas terrestres, sobre todo en forma de lignina. Otros compuestos aromáticos no son producidos de forma natural, sino por la actividad de las industrias farmacéutica, química, agrícola o alimentaria. Además, los compuestos conocidos como BTEX (acrónimo de benceno, tolueno, etilbenzeno y xileno) son contaminantes ambientales comunes procedentes de derivados del petróleo (Fuchs y col., 2011).

La acumulación más preocupante de compuestos aromáticos en el medio se debe principalmente a la actividad humana y pueden ocasionar un importante daño ambiental. Una gran cantidad de compuestos orgánicos industriales son liberados al medio como herbicidas y pesticidas, otros se vierten accidentalmente o como residuos de las distintas industrias. La persistencia de estos compuestos en los ecosistemas puede ser de semanas, si se

## **Introducción**

descomponen espontáneamente en condiciones naturales o son degradados fácilmente por los microorganismos, o de muchos años si son difícilmente biodegradables (Dua y col., 2002). Las razones por las cuales los compuestos aromáticos son muy persistentes en el medio son su limitada reactividad química, su hidrofobicidad, que dificulta su biodisponibilidad, y, en el caso de los nuevos compuestos industriales, la falta de adaptación de los microorganismos para poder degradarlos (Pieper y Reineke, 2000; Fuchs y col., 2011).

A pesar de las dificultades para eliminar los compuestos aromáticos del medio, existen microorganismos que han desarrollado la capacidad de utilizarlos de distintas formas. Algunos pueden usarlos en anaerobiosis como aceptores de electrones, como ocurre en la dehalorespiración, en la que los compuestos aromáticos clorados actúan como aceptores de electrones o como intermediarios en la cadena de transporte de electrones, aunque en ambos casos el anillo bencénico queda inalterado y los cambios ocurren en los radicales (Gibson y Harwood, 2002). Sin embargo, es el uso de estos compuestos como fuente de carbono y energía, tanto en condiciones aerobias como anaerobias, lo que permite transformarlos en compuestos menos tóxicos, inocuos o incluso mineralizarlos completamente a dióxido de carbono y agua. A pesar de ello, hay que tener en cuenta que la degradación parcial de algunos compuestos aromáticos puede generar intermediarios incluso más tóxicos que la molécula inicial, por lo que los microorganismos capaces de mineralizarlos son los verdaderamente importantes desde un punto de vista ambiental (Ghosal y col., 2016).

En la degradación de compuestos aromáticos, los genes catabólicos suelen agruparse en operones. Junto a ellos, frecuentemente se encuentran los genes que codifican para el transportador del metabolito aromático y los reguladores transcripcionales específicos de la ruta. Estos factores específicos pertenecen a un amplio espectro de familias distintas de reguladores que han evolucionado para controlar la expresión de los operones catabólicos de un compuesto en concreto, lo cual asegura la expresión de las enzimas catabólicas y de los transportadores solo en el caso de que el compuesto esté presente en el medio, es decir, en condiciones de inducción. Además, el catabolismo de los compuestos aromáticos no está sujeto exclusivamente a una regulación específica, sino también a una regulación global que implica una respuesta metabólica que conecta la ruta catabólica específica del compuesto aromático con el metabolismo energético de la célula y una respuesta frente al estrés que causa el carácter tóxico de la molécula aromática (Díaz y col., 2013).

Dentro de la respuesta global, el catabolismo de los compuestos aromáticos se suele ver afectado por el fenómeno de represión catabólica por carbono (CCR, acrónimo en inglés) por el que se establece una prioridad de consumo de fuentes de carbono si existe una mezcla de los mismos en el medio. La CCR mejora la habilidad de las bacterias para competir en sus

hábitats naturales a través de la optimización de su metabolismo, puesto que consumen primero la fuente de carbono preferencial, es decir, la que pueden metabolizar más rápidamente y con la que obtienen un crecimiento más eficiente. Para conseguir lo anterior, existen distintos mecanismos moleculares por los cuales las bacterias inhiben tanto el transporte como la expresión de los genes requeridos para el catabolismo de los compuestos no preferenciales (Rojo, 2010). En general, el fenómeno de CCR puede llegar a alterar la expresión de hasta el 5-10 % de los genes de un genoma bacteriano (Blencke y col., 2003; Liu y col., 2005).

Los compuestos aromáticos constituyen fuentes de carbono secundarias para las bacterias, por lo que la CCR representa una barrera a su degradación en un medio complejo en el que estén presentes otros metabolitos preferenciales, como suelen ser azúcares, ácidos orgánicos o aminoácidos. Como consecuencia, el estudio de las bases moleculares que rigen la regulación de la CCR en bacterias degradadoras de aromáticos resulta imprescindible para optimizar las estrategias de eliminación de estos compuestos del medio (Rojo, 2010). A pesar de su relevancia, la represión catabólica sólo se ha estudiado extensivamente en las bacterias *Escherichia coli*, *Pseudomonas putida* y *Bacillus subtilis*, encontrándose mecanismos moleculares diferentes en cada una de ellas (revisado en Görke y Stülke, 2008).

## **2. La familia *Sphingomonadaceae***

La familia *Sphingomonadaceae* contiene bacterias capaces de degradar una gran diversidad de compuestos recalcitrantes. Junto con la familia *Erythrobacteraceae*, constituyen el orden *Sphingomonadales*, el cual está englobado dentro de la clase  $\alpha$ -proteobacteria (Yabuuchi y Kosako, 2015). Los miembros de la familia *Sphingomonadaceae* son bacterias Gram-negativas de distintos tamaños que no forman esporas, pueden ser móviles o no y, en el caso de serlo, suelen presentar un flagelo polar. Las colonias que forman presentan tonos amarillos o anaranjados por la presencia de carotenoides y contienen esfingolípidos (glicoesfingolípidos) en sus envueltas celulares en lugar de lipopolisacáridos. Otras características de la familia son la presencia de la ubiquinona 10 como la principal quinona respiratoria, del ácido graso octadecenoico (C18:1) como el principal ácido graso del total de lípidos extraíbles y la predominancia de las poliaminas *sym*-espermidina y espermidina (revisado en Glaeser y Kämpfer, 2014).

Dentro de la familia *Sphingomonadaceae* se ha definido el grupo de las esfingomonas constituido por los géneros *Sphingomonas*, *Sphingobium*, *Novosphingobium* y *Sphingopyxis* (Takeuchi y col., 2001). Este grupo de bacterias se encuentra ampliamente distribuido en la naturaleza, desde distintos ambientes terrestres y acuáticos, hasta raíces de plantas,

## Introducción

muestras clínicas y otras fuentes (Balkwill y col., 2006). Algunas cepas se han aislado incluso de relaves mineros contaminados con arsénico (Macur y col., 2001) y de lodos activos procedentes de las plantas de tratamiento de aguas residuales (Hu y col., 2011). Además, este grupo comprende bacterias metabólicamente muy versátiles que presentan la capacidad de utilizar como fuente de carbono y energía un gran número de compuestos recalcitrantes, tanto naturales como sintéticos, principalmente metabolitos aromáticos y sus derivados (Balkwill y col., 2006).

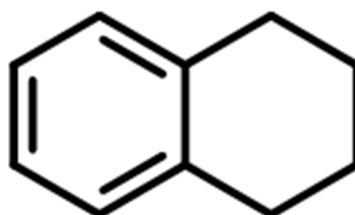
Entre los metabolitos que son capaces de degradar las bacterias del grupo esfingomonas se encuentran los PAHs, compuestos aromáticos derivados de la lignina, compuestos aromáticos clorados y sulfonados, fenoles e incluso pesticidas y herbicidas aromáticos (Stolz, 2009). Algunos ejemplos de compuestos aromáticos degradados por las bacterias del grupo de las esfingomonas son el bifenilo, naftaleno, *m*-xileno y el *p*-cresol, todos catabolizados por la bacteria *Novosphingobium* (antes *Sphingomonas*) *aromaticivorans* F199 (Romine y col., 1999); la dibenzo-*p*-dioxina, degradada por *Sphingomonas wittichii* RW1 (Wittich y col., 1992); el pentaclorofenol, utilizado como fuente de carbono y energía por *Sphingobium chlorophenolicum* L-1 (Copley y col., 2012) o la tetralina, catabolizada por *Sphingopyxis granuli* TFA (López-Sánchez, 2010 y referencias en él), que es la bacteria objeto de estudio en esta tesis doctoral (anteriormente clasificada como *Sphingomonas macrogoltabida*).

Los compuestos aromáticos no son las únicas moléculas contaminantes que son capaces de degradar las bacterias del grupo de las esfingomonas. Se ha descrito la capacidad de este grupo para utilizar polímeros como el polietilenglicol (PEG) o el alcohol de polivinilo (PVA) (Kaway, 1999). Ambos son polímeros sintéticos solubles en agua que se utilizan en la preparación de artículos de plásticos, adhesivos, films para embalajes, o como aditivos en las industrias papelera y maderera, en pinturas y textiles (Chiellini y col., 2003; Kawai y col., 2013). Se trata de polímeros biodegradables que, aunque no son tóxicos para los organismos, pueden producir grandes cantidades de espuma en los sistemas acuáticos, lo cual disminuye la disponibilidad de oxígeno en el agua (Kawai y col., 2013). Por tanto, la degradación de estos polímeros, tanto microbiana como enzimática, ha suscitado un gran interés como forma de eliminarlos del agua. Dentro de las bacterias del grupo de las esfingomonas, se han aislado cepas capaces de degradar PVA, como *Sphingopyxis* sp. 113P3 (Hu y col., 2008) y *Sphingopyxis* sp. PVA3 (Yamatsu y col., 2006), y otras con la habilidad de utilizar PEG, como *Sphingopyxis macrogoltabida* y *Sphingopyxis terrae* (Takeuchi y col., 2001). Se ha descrito que estas bacterias despolimerizan el PVA y el PEG en el periplasma, aunque se desconoce el mecanismo de entrada a dicho compartimento (Kawai y col., 2013).

Normalmente, los genes codificantes para las enzimas implicadas en la degradación de xenobióticos en las bacterias del grupo de las esfingomonas se encuentran en mega-plásmidos y se organizan en varias unidades transcripcionales e incluso en diferentes replicones (Stolz, 2014). Por ejemplo, todos los genes implicados en la degradación de aromáticos en *Novosphingobium aromaticivorans* F199, se localizan en el mega-plásmido pNL1 (Romine y col., 1999; Stolz, 2014). Las enzimas para la degradación de PVA en *Sphingopyxis* sp. 113P3 también se localizan en un mega-plásmido (Hu y col., 2008). En el caso de *Sphingobium japonicum* UT26, los genes para la degradación de  $\gamma$ -hexaclorociclohexano se organizan en tres replicones distintos, incluidos los dos cromosomas que se han definido en esta bacteria y un plásmido (Nagata y col., 2006). Esta organización génica flexible y el mantenimiento y diseminación de los mega-plásmidos, sugieren que los fenómenos de transferencia horizontal han contribuido en gran medida a que este grupo de bacterias se adapte de forma rápida a la degradación de distintos compuestos xenobióticos y contenga una elevada representación de bacterias degradadoras (Stolz, 2014; Ghosal y col., 2016).

### 3. Degradación de tetralina en *S. granuli* TFA

Como ya se ha mencionado anteriormente, la bacteria objeto de estudio en esta tesis doctoral es *Sphingopyxis granuli* TFA, perteneciente a la familia *Sphingomonadaceae* y al grupo de las esfingomonas. Esta bacteria Gram-negativa presenta la capacidad de utilizar el solvente orgánico tetralina (1,2,3,4-tetrahidronaftaleno) como única fuente de carbono y energía (Hernández y col., 1999). La molécula de tetralina está formada por un anillo aromático y otro alicíclico que comparten dos átomos de carbono (Figura 1).



**Figura 1. Molécula de tetralina.** A la izquierda se representa el anillo aromático y a la derecha el anillo alicíclico.

## Introducción

Se trata de un compuesto producido industrialmente mediante la hidrogenación del naftaleno o el craqueo del antraceno el cual se utiliza como sustituyente del aguarrás en pinturas y ceras o en la industria petroquímica como solvente en el proceso de licuefacción del carbón. La tetralina se encuentra también en el alquitrán y el petróleo (Sikkema y Bont, 1993).

Debido a su carácter lipofílico, la molécula de tetralina puede interaccionar con las membranas biológicas provocando cambios en su estructura y función, lo que llevaría a un mal funcionamiento de la actividad celular. Además de provocar un incremento en la permeabilidad de las membranas, la tetralina puede producir hidroperóxidos, altamente mutagénicos, en el interior celular (Ferrante y col., 1995). Se ha determinado que concentraciones de tetralina superiores a 100  $\mu\text{M}$  son tóxicas para cultivos microbianos (Sikkema y Bont, 1991) y, aprovechando esta propiedad, se ha utilizado como biocida (Sikkema y col., 1992).

### 3.1. La ruta de degradación de tetralina

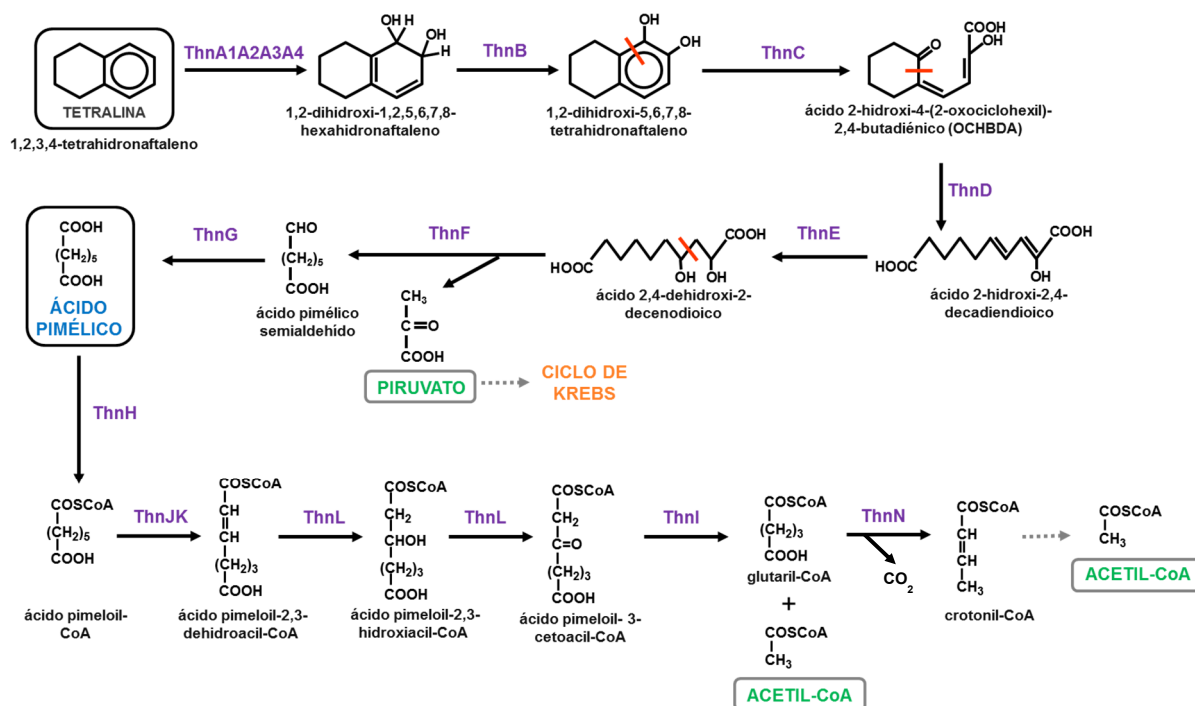
Se han aislado algunas bacterias capaces de utilizar la tetralina como única fuente de carbono y energía (Sikkema y Bont, 1991) aunque la ruta de degradación no se ha estudiado en profundidad en la mayoría de ellas. En *Corynebacterium sp.* estirpe C125 sólo se han caracterizado los pasos iniciales de la ruta de degradación que corresponden con la rotura del anillo aromático (Sikkema y Bont, 1993). Del mismo modo, en *Pseudomonas stutzari* AS39 únicamente se han descrito los primeros pasos de degradación, que en este caso consisten en una hidroxilación y posterior oxidación del anillo alicíclico (Schreiber y col., 1983).

De todas las bacterias que degradan tetralina, la ruta de degradación sólo ha sido estudiada y caracterizada al completo en *Sphingopyxis granulii* TFA. Dicha caracterización ha sido posible gracias a la construcción de mutantes, secuenciación génica e identificación de metabolitos intermediarios de la ruta (Hernández y col., 1999; Andújar y col., 2000; Hernández y col., 2000; Hernández y col., 2002; Andújar y Santero, 2003; Moreno-Ruiz y col., 2003; López-Sánchez y col., 2010). Además, esta bacteria crece en tetralina como única fuente de carbono y energía con un tiempo de generación de 12,5 horas, siendo el mejor de los descritos hasta el momento (Sikkema y Bont, 1991).

Como se muestra en la Figura 2, la degradación de tetralina en TFA comienza con la rotura en la posición *meta* del anillo aromático, lo cual ocurre mediante una dioxigenación inicial (catalizada por el complejo enzimático ThnA1A2A3A4) y una posterior deshidrogenación (catalizada por la enzima ThnB) (Moreno-Ruiz y col., 2003). El producto obtenido tras estas reacciones es 1,2-dihidroxitetralina, el cual constituye el sustrato para la



dioxigenasa ThnC que produce la rotura extradiólica del derivado catecólico (Andújar y col., 2000).



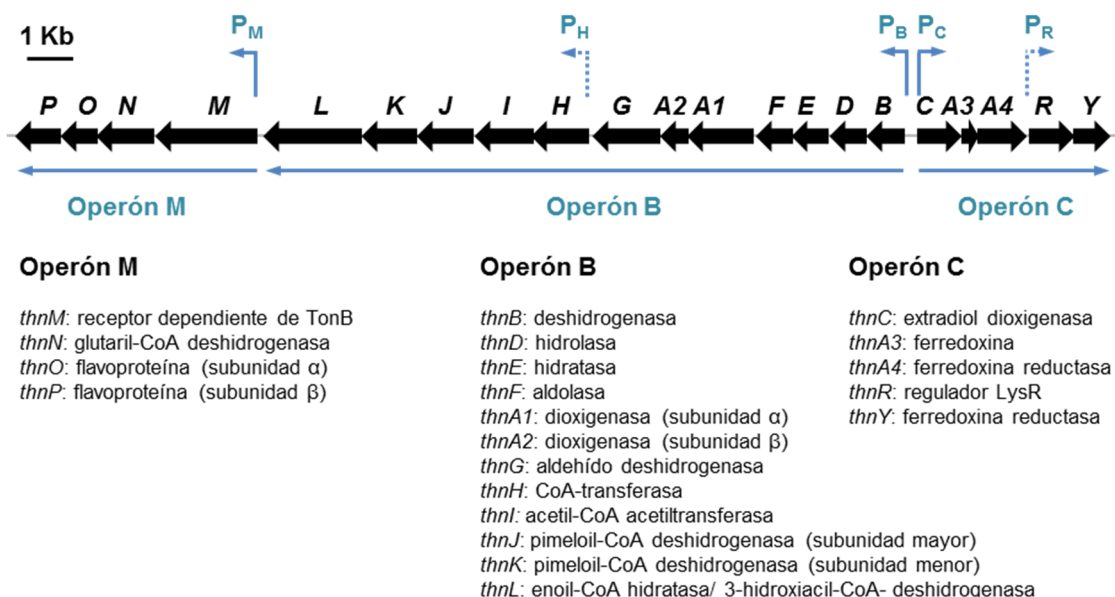
**Figura 2. Ruta de degradación de tetralina en *S. granuli* TFA.** Las rayas rojas en los compuestos señalan el enlace que se rompe y las flechas discontinuas indican que hay más de una reacción.

Una vez abierto el anillo aromático, la hidrolasa ThnD realiza la hidrólisis del enlace C-C del anillo alicíclico, generándose un ácido dicarboxílico largo de 10 átomos de carbono (Hernández y col., 2000). A partir de este compuesto, la acción secuencial de una hidratasa (ThnE) y una aldolasa (ThnF) termina en la formación de una molécula de ácido pimélico semialdehído y otra de piruvato (Hernández y col., 2002). Mientras que el piruvato entra directamente al metabolismo central a través del ciclo de Krebs, el ácido pimélico semialdehído se oxida a través de una aldehído deshidrogenasa (ThnG) hasta ácido pimélico, un ácido dicarboxílico de siete carbonos que se metaboliza mediante  $\beta$ -oxidación por la acción secuencial de las enzimas ThnH, ThnJ, ThnK, ThnL, ThnI y ThnN, hasta formar una molécula de acetil-CoA y otra de crotonil-CoA. Finalmente, la molécula de crotonil-CoA puede seguir metabolizándose hasta formar otra de acetil-CoA (López-Sánchez y col., 2010).

Las proteínas ThnO y ThnP (subunidades  $\alpha$  y  $\beta$  de flavoproteína), no representadas en la Figura 2, son las encargadas de aceptar los electrones de las deshidrogenasas de la ruta de  $\beta$ -oxidación, los cuales son posteriormente transferidos a la cadena respiratoria.

## Introducción

Los genes reguladores y estructurales implicados en la degradación de tetralina en *S. granuli* TFA, denominados genes *thn*, se encuentran agrupados en tres operones distintos (operón B, C y M), los cuales se expresan desde sus respectivos promotores  $P_B$ ,  $P_C$  y  $P_M$ . Además existen dos operones internos,  $P_H$  y  $P_R$ , situados en el operón B y C respectivamente (López-Sánchez y col., 2009; López-Sánchez y col., 2010) (Figura 3).



**Figura 3. Organización génica de los genes implicados en la degradación de tetralina en *S. granuli* TFA.** Con flechas azules se delimitan los operones y se indica la posición de los promotores y su direccionalidad.

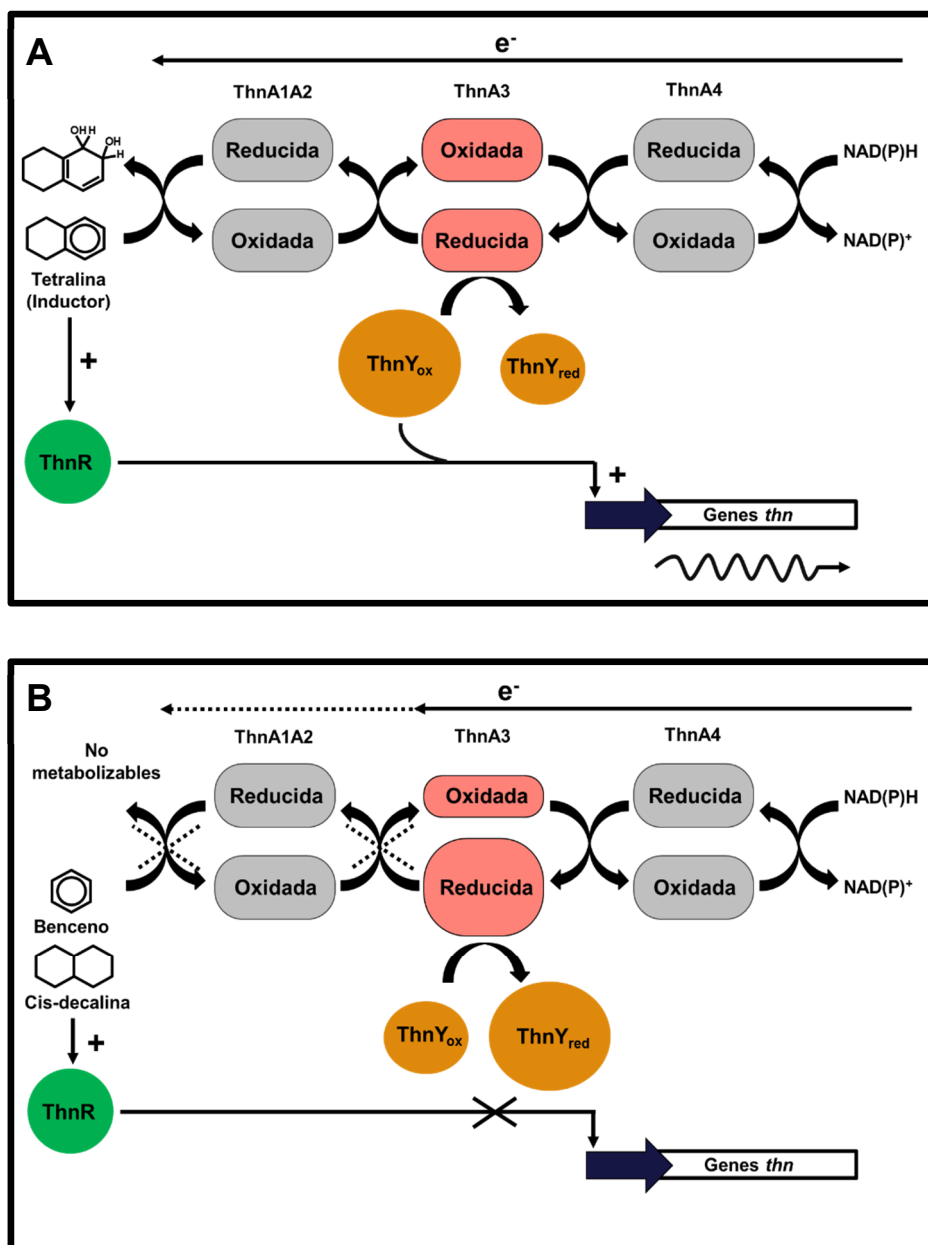
Los genes imprescindibles para la metabolización de la tetralina son todos los contenidos en el operón C y los genes desde *thnB* hasta *thnA2* contenidos en el operón B (Hernández y col., 1999; Moreno-Ruiz y col., 2003; Martínez-Pérez y col., 2004; López-Sánchez y col., 2010). El resto de genes no son imprescindibles puesto que en el genoma de TFA se encuentran codificados otros operones para llevar a cabo la  $\beta$ -oxidación de ácidos grasos que harían la misma función.

### 3.2. Regulación específica de los genes *thn* de *S. granuli* TFA

Se sabe que la expresión de los operones de degradación de tetralina se regula a nivel transcripcional y que los elementos imprescindibles para ello son la molécula de tetralina (inductor), el activador transcripcional tipo LysR, ThnR, y su coactivador, ThnY (Martínez-Pérez y col., 2004; García y col., 2011). Además, la ferredoxina ThnA3, que transfiere electrones a la dioxigenasa de tetralina (ThnA1A2), juega un papel importante en la transducción de la señal (Martínez-Pérez y col., 2007). Mientras que ThnR y ThnY pueden

activar de forma promiscua a los promotores  $P_B$ ,  $P_C$ ,  $P_H$  y  $P_M$ , ThnA3, en comunicación con ThnY, se encarga de evitar la activación de la transcripción cuando la molécula inductora no es metabolizable a través de la ruta de degradación (Ledesma-García y col., 2013).

El gen *thnR* se expresa de forma constitutiva desde el promotor interno  $P_R$  incluso en ausencia de inductor y su expresión aumenta en presencia de tetralina al transcribirse desde el promotor  $P_C$  junto con el resto de genes contenidos en el operón C (Martínez-Pérez y col., 2004; López-Sánchez y col., 2009). Para activar la expresión de los genes de la ruta de degradación de tetralina, ThnR necesita, además de la molécula de tetralina, al coactivador ThnY, el cual, al igual que ThnR, se transcribe constitutivamente desde el promotor  $P_R$  y de forma inducida desde el promotor  $P_C$ . ThnY es una proteína homóloga a las ferredoxinas reductasas que transportan electrones desde el NAD(P)H hasta las dioxigenasas terminales implicadas en la degradación de compuestos aromáticos. Sin embargo, se ha comprobado que el papel de ThnY en TFA no es catalítico sino regulador y que ocurre a través de una comunicación con el complejo de la dioxigenasa de tetralina (ThnA1A2A3A4) (Martínez-Pérez y col., 2007; García y col., 2011). Por último, se ha demostrado que esta comunicación ocurre a través de una transferencia secuencial de electrones desde el NAD(P)H hasta el coactivador ThnY, la cual ocurre vía ThnA4 y ThnA3. Cuando la molécula inductora no es sustrato de la dioxigenasa inicial, la transferencia de electrones se produce hacia el coactivador ThnY que se acumula en su forma reducida (inactiva) con lo que se evita la inducción de los genes *thn* (Ledesma-García y col., 2013; Ledema-García y col., 2016). El modelo de regulación de los genes *thn* considerando los elementos anteriores se ilustra en la Figura 4.



**Figura 4. Modelo de regulación específica de la expresión de los genes *thn* en presencia de la molécula inductora (A) o sustratos no metabolizables (B).** El tamaño de los círculos representa la cantidad de proteína en cada uno de los estados y las líneas punteadas el bloqueo en la transferencia de electrones. Ni el benceno ni la cis-decalina son sustratos de la dioxigenasa inicial. Imagen adaptada de Ledesma-García y col. (2016).

### 3.3. Represión catabólica de los genes *thn* de *S. granuli* TFA

Además de la regulación específica, los genes de degradación de tetralina están sujetos a represión catabólica por carbono (CCR). En *S. granuli* TFA, la presencia en el medio de cultivo de una fuente de carbono preferencial impide la inducción por tetralina de los genes *thn* (Martínez-Pérez y col., 2004).

Se sabe que *S. granuli* TFA utiliza un rango reducido de fuentes de carbono. En contra de lo que sucede en la mayoría de las bacterias, TFA no es capaz de crecer en azúcares simples, como la glucosa, ni en disacáridos. Tampoco ha mostrado crecimiento en ácidos grasos mono o dicarboxílicos de cadena corta, como el succinato, el 2-oxoglutarato o el piruvato. Antes de la realización de esta tesis doctoral, se había comprobado que TFA crece en acetato,  $\beta$ -hidroxibutirato, ácidos grasos de cadena media-larga como el ácido sebácico (10 C), el ácido mirístico (14 C) y el ácido tetradecanodioico (14 C). En cuanto a los aminoácidos, TFA es capaz de crecer en prolina y glutamina, los cuales sirven también como fuente de nitrógeno. De todas estas fuentes de carbono, el  $\beta$ -hidroxibutirato a una concentración de 40 mM y el ácido sebácico a 16 mM son en las que que TFA se multiplica con un menor tiempo de generación (Hernández Silva, 2000).

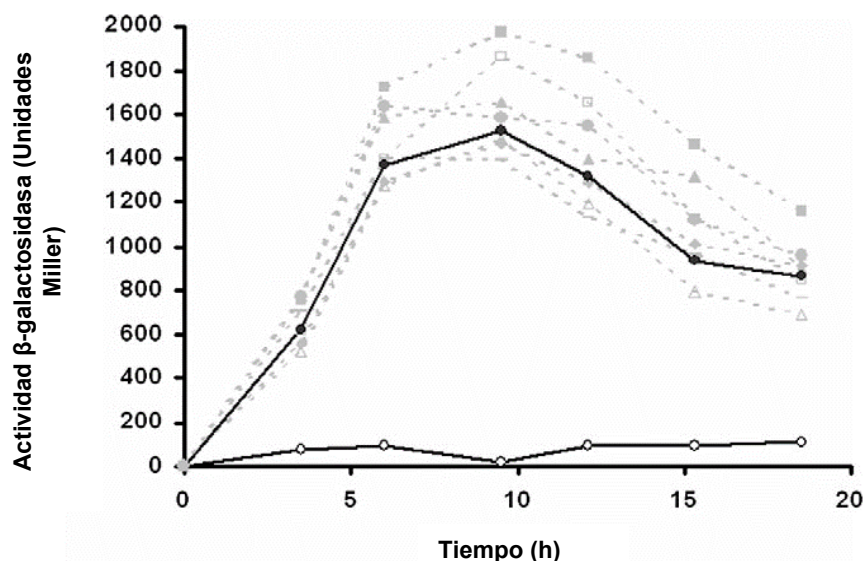
Se ha demostrado que, cuando en el medio existe una fuente preferencial de carbono, los genes *thn* solo comienzan a inducirse cuando esta comienza a agotarse (pasadas las 10 horas en MML -medio rico- más tetralina y pasadas las 15 horas en  $\beta$ -hidroxibutirato 40 mM más tetralina). Además, la magnitud de la CCR es proporcional a la concentración de la fuente de carbono preferencial, afectando tanto a los niveles de inducción de los genes *thn* como al tiempo al que se inicia la inducción, de forma que a mayor concentración, menor nivel de inducción y mayor tiempo de respuesta (Martínez-Pérez y col., 2004).

Conocer los mecanismos moleculares por los cuales tiene lugar la represión catabólica de los genes *thn* en *S. granuli* TFA es un paso importante en el entendimiento de la regulación de la ruta de degradación de tetralina en esta bacteria. Además, este fenómeno no se ha caracterizado en ninguna bacteria perteneciente al grupo de las esfingomonas, a pesar de ser un grupo muy importante en la biodegradación de compuestos recalcitrantes como ya se comentó anteriormente. Tampoco se conocen las bases moleculares de la represión catabólica en la clase  $\alpha$ -proteobacteria, a excepción de *Sinorhizobium meliloti* donde se ha estudiado con algo más de detalle (Pinedo y col., 2008; Pinedo y Gage, 2009; García y col., 2010).

Para detectar elementos implicados en el fenómeno de CCR en TFA, se había realizado previamente en el laboratorio una mutagénesis por transposición de miniTn5-Km (Martín-

## Introducción

Cabello y col., 2011). De esta forma se consiguieron aislar hasta 8 mutantes que, en condiciones de represión catabólica, expresaban los genes *thn* (Figura 5).



**Figura 5. Expresión de la fusión *thnC::lacZ* en TFA silvestre y en los ocho candidatos de la mutagénesis con transposón.** La actividad  $\beta$ -galactosidasa se midió en células creciendo en medio mínimo más tetralina y  $\beta$ -hidroxibutirato 40 mM. Dentro de los mutantes (líneas punteadas) se resalta la estirpe MPO209 (círculos negros). La estirpe silvestre se representa con círculos blancos. Imagen adaptada de Martín-Cabello y col. (2011).

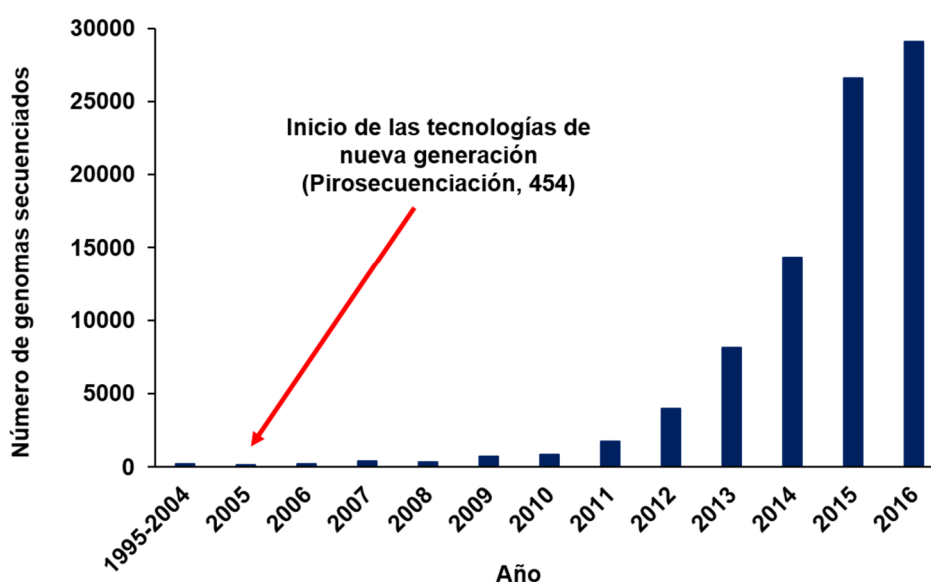
En todos los mutantes la inserción había ocurrido en un gen codificante para una polihidroxicanoato sintasa (*phaC*) o en el gen situado justo aguas arriba, causando probablemente un efecto polar. El gen *phaC* está implicado en la síntesis de gránulos de polihidroxibutirato (PHB) y su ausencia reduce la producción de dicho polímero (Martín-Cabello, 2011). La desrepresión de los genes *thn* observada en los mutantes anteriores es parcial y, por tanto, debe haber otros mecanismos adicionales que regulen el fenómeno de CCR de los mismos.

## 4. Secuenciación, ensamblaje y anotación de los genomas bacterianos

En el año 2012, al inicio de esta tesis doctoral, sólo se habían secuenciado 26 genomas de bacterias del grupo de las esfingomonas, sólo 7 estaban almacenados como genomas completos en GenBank y el único genoma secuenciado del género *Sphingopyxis* pertenecía a la bacteria *Sphingopyxis alaskensis* RB2256 (Fuente: fichero prokaryotes.txt de GenBank). La escasa representación de genomas completamente secuenciados del grupo de las esfingomonas y la importancia de este grupo en la degradación de compuestos recalcitrantes ponen en valor el desarrollo de proyectos de secuenciación y anotación de bacterias

pertenecientes a este grupo, tanto para estudiar su potencial genético y metabólico como para ampliar el conocimiento sobre las rutas de degradación de contaminantes y sus mecanismos de regulación específicos y globales.

Han pasado más de dos décadas desde que se secuenciara el primer genoma procariota, perteneciente a la bacteria Gram-negativa *Haemophilus influenzae* (Fleischmann y col., 1995). Desde entonces, los avances en las tecnologías de secuenciación y el desarrollo de herramientas bioinformáticas han hecho posible una significativa reducción de los costes y del tiempo requerido para secuenciar un genoma completo, permitiendo que muchos laboratorios puedan afrontar la secuenciación de sus organismos de interés (Land y col., 2015). Como consecuencia, el número de genomas procariotas secuenciados desde 1995 se ha ido incrementando con una tendencia exponencial (Figura 6), sobre todo tras la primera década, coincidiendo con el desarrollo de las tecnologías de secuenciación masiva de nueva generación (*Next-Generation Sequencing*, NGS).



**Figura 6. Número de genomas de bacterias y arqueas secuenciados cada año y subidos a NCBI.** Fuente: fichero prokaryotes.txt de GenBank descargado el 19 de Julio de 2017. Gráfica actualizada de la revisión de Land y col., 2015.

En la actualidad, el gran número de genomas bacterianos secuenciados y almacenados en las bases de datos permite que la secuenciación de un genoma no sólo sea de utilidad para la caracterización de dicho organismo, mediante la reconstrucción de su metabolismo o el análisis de la expresión génica utilizando experimentos de transcriptómica global, sino que también posibilita la realización de análisis taxonómicos a escala genómica y distintos estudios de genómica comparada.

### 4.1. Tecnologías para la secuenciación de genomas

Las primeras estrategias de secuenciación surgieron hace unos 40 años. En la década de 1970, Frederick Sanger y sus colaboradores desarrollaron la secuenciación Sanger (Sanger y Coulson, 1975; Sanger y col., 1977), lo cual supuso una revolución en el campo de la biología y permitió la secuenciación del primer genoma de un organismo de vida libre que, como se mencionó anteriormente, fue el de la bacteria *Haemophilus influenzae* en 1995. Sin embargo, hasta principios del siglo XXI no comenzaron a desarrollarse las primeras tecnologías de secuenciación masiva de nueva generación que desbancaron a la secuenciación Sanger como metodología para la obtención de la secuencia de genomas completos. Existen diversas tecnologías diferentes de secuenciación masiva, como son la pirosecuenciación 454 (el utilizado para la secuenciación del genoma de la bacteria objeto de esta Tesis, aunque ya en desuso), Illumina, SOLiD y la secuenciación SMRT (*Single Molecule Real Time*) (revisado en Pareek y col., 2011 y en Loman y Pallen, 2015).

### 4.2. Ensamblado de las secuencias procedentes de las tecnologías de secuenciación masiva

El resultado de la secuenciación masiva consiste en una gran cantidad de pequeñas secuencias de ADN, conocidas como lecturas. Dependiendo de la tecnología de secuenciación empleada, estas lecturas varían en tamaño y cantidad y son proporcionadas por las empresas que ofrecen el servicio, con información sobre la calidad de la secuenciación. Los secuenciadores 454 de Roche GS-FLX Titanium generan lecturas que rondan las 400 pares de bases (Miller y col., 2010) y como resultado devuelven un fichero en formato SFF (*Standard Flowgram Format*). Este formato contiene información sobre la calidad con la que ha sido secuenciado cada nucleótido de las distintas lecturas, además de otra información relevante del proceso de secuenciación (Malde, 2011). A partir del fichero SFF se puede obtener un fichero FASTA que contenga las secuencias de todas las lecturas, ya que éste es el formato requerido como entrada para multitud de herramientas bioinformáticas.

Para obtener la secuencia completa de un genoma se requiere el ensamblaje de las pequeñas lecturas de la secuenciación. El proceso de ensamblaje consiste en alinear y unir las pequeñas lecturas en fragmentos de mayor tamaño para obtener la secuencia ordenada del genoma. Se han desarrollado multitud de algoritmos bioinformáticos para el ensamblado de las secuencias, denominados ensambladores (revisado en Miller y col., 2010). Existen dos procedimientos diferentes de ensamblaje según el caso de estudio: el ensamblaje *de novo* y el ensamblaje mediante mapeo en un genoma de referencia. En el primer caso, no se dispone



de información previa sobre el genoma a secuenciar, mientras que en el segundo, las pequeñas lecturas se ensamblan utilizando una secuencia molde como referencia.

Entre los distintos ensambladores *de novo* se encuentran Newbler (454 Life Sciences), diseñado específicamente para el ensamblado de secuencias largas procedentes de las plataformas de pirosecuenciación 454, o Celera Assembler (Myers y col., 2000), utilizado para el ensamblaje de secuencias 454 y de secuencias generadas por otras plataformas de secuenciación como Illumina.

En la mayoría de los casos de ensamblaje *de novo*, no puede cerrarse la secuencia del genoma completo mediante el procedimiento *in silico*. En su lugar, se obtienen secuencias ensambladas de forma contigua que finalizan en regiones donde los ensambladores automáticos no pueden resolver la secuencia (Kisand y Lettieri, 2013). Para terminar de cerrar un genoma, se deben ordenar los fragmentos generados por los ensambladores automáticos y obtener la secuencia de los huecos entre ellos. Este paso constituye el cuello de botella en el proceso de ensamblaje y, generalmente, requiere de más intervención manual que las etapas anteriores.

Para evaluar la calidad del ensamblaje y analizar la cobertura del genoma por las lecturas de la secuenciación, éstas se alinean con los fragmentos obtenidos o, en su caso, con el genoma completo. De esta forma pueden detectarse errores en el ensamblaje y otros artefactos que faciliten el refinamiento final de la secuencia. Para este fin existen distintos algoritmos de alineamiento de lecturas frente a una secuencia de referencia, como BWA (*Burrows-Wheeler Aligner*) (Li y Durbin, 2009).

#### **4.3. Anotación y caracterización de la información genómica**

Tras finalizar el procedimiento de secuenciación y ensamblaje se obtiene un genoma completo o parcialmente ensamblado. El siguiente paso es realizar una anotación de la secuencia primaria disponible para definir los elementos que componen el genoma y asignarles información biológica. El conocimiento de la presencia y localización de dichos elementos junto con el estudio de sus funciones son necesarios para revelar los procesos biológicos que tienen lugar en el organismo así como su historia evolutiva (Nawrocki, 2014). De forma análoga al proceso de secuenciación, la anotación de un genoma completo es un procedimiento muy automatizado (Médigue y Moszer, 2007). En el proceso de anotación, inicialmente se delimitan los distintos elementos en el genoma, lo cual se conoce como anotación estructural, y después se les da una función, lo que constituye la anotación funcional. Finalmente, la anotación de los elementos del genoma debe organizarse en un

fichero de forma ordenada para acceder fácilmente a la información (Stothard y Wishart, 2006). Para las distintas partes del proceso de anotación (anotación estructural y funcional) existen distintas herramientas bioinformáticas especializadas que se describen a continuación.

### 4.3.1. Anotación estructural de genes codificantes de proteínas

La anotación estructural de genes codificantes de proteínas a lo largo de la secuencia genómica consiste en la identificación de las pautas abiertas de lectura (ORFs acrónimo en inglés), definiendo el codón de inicio y fin para cada gen codificante. Esta identificación se realiza tanto en la cadena de ADN que se utiliza en el análisis como en su reversa complementaria. Existen distintas herramientas bioinformáticas para la detección de ORFs en genomas procariotas, entre las cuales destacan GeneMarkS (Besemer y col., 2001), Glimmer (Salzberg y col., 1998; Delcher y col., 2007) y Prodigal (Hyatt y col., 2010). A pesar de que estas herramientas permiten detectar las ORFs de un genoma procariota completo en unos pocos minutos, presentan algunos problemas para determinar el codón de inicio correcto y pueden generar falsas predicciones. La dificultad para identificar el inicio de la región codificante radica en la falta de elementos en las secuencias que definan de forma clara qué codón de los posibles es el iniciador (Besemer y col., 2001). En bacterias, los codones de inicio pueden ser AUG, GUG y UUG (McCarthy y col., 1994) y, excepcionalmente, AUU, identificado como el codón de inicio para el factor de transcripción IF3, el cual está codificado por el gen *infC* de *Escherichia coli* (McCarthy y col., 1994), y para una poliA-polimerasa codificada por el gen *pcnB* en *E. coli* (Binns y Masters., 2002). Los porcentajes de uso de los codones de inicio AUG, GUG y UUG en *E. coli* son del 90, 8 y 1 %, respectivamente (Laursen y col., 2005), y parecen ser similares en la mayoría de genomas bacterianos.

La mayoría de predictores de ORFs consideran la secuencia de unión al ribosoma (RBS, acrónimo en inglés, o secuencia Shine-Dalgarno, SD) para definir el codón de inicio con mayor precisión. Sin embargo, la diversidad de RBS en los genomas bacterianos e incluso la expresión de genes a los cuales les falta una secuencia SD hacen que esta estrategia, a pesar de ser una importante aproximación, no evite la aparición de falsos positivos o la pérdida de predicción de algunas ORFs (Omotajo y col., 2015). Aunque existen dificultades para definir con precisión el inicio de la traducción de las proteínas codificadas en un genoma y evitar falsas predicciones, la herramienta Prodigal ha sido diseñada para mejorar la predicción de las ORFs, mejorar el reconocimiento del sitio de iniciación de la traducción y reducir el número de genes predichos como falsos positivos. Además, ha mostrado ser el método de predicción más robusto en varios genomas bacterianos (Hyatt y col., 2010; Omotajo y col., 2015).

#### 4.3.2. Anotación funcional de genes codificantes de proteínas

La anotación funcional de las proteínas definidas previamente consiste en la asignación de información biológica a cada una de ellas. A pesar de que algunas secuencias se anotan usando evidencias experimentales directas, la mayoría de anotaciones se realizan *in silico* y se basan en la búsqueda de similitud con las proteínas almacenadas en las bases de datos (Loewenstein y col., 2009). En general, si en esta comparación se obtiene un resultado con una cierta similitud, la información se transfiere a la nueva proteína. Además de proporcionar una función, en los procedimientos de anotación se puede incluir otro tipo de información de la proteína, como sus propiedades químicas y estructurales, la ontología o la ruta metabólica en la cual está implicada (Stothard y Wishart, 2006).

A pesar de que idealmente la anotación automática es un proceso rápido y sencillo para conseguir obtener la función de todas las proteínas de un genoma, no todas las secuencias que se quieren anotar presentan similitud con las almacenadas en las bases de datos y, por tanto, representarían nuevas proteínas no caracterizadas cuya anotación requeriría de procedimientos experimentales, comparación de similitud a nivel de estructura secundaria o alguna otra estrategia, como la búsqueda de posibles dominios conservados. Para realizar la anotación funcional se han desarrollado numerosas herramientas bioinformáticas, algunas de ellas basadas en servidores web como son RAST (Aziz y col., 2008) o BASys (Van Domselaar y col., 2005) y otras que se ejecutan de forma local, como los programas Prokka (Seemann, 2014) o Sma3s (Muñoz-Mérida y col., 2014; Casimiro-Soriger y col., 2017), los cuales están implementados en lenguaje de programación Perl.

La herramienta RAST es un servidor web diseñado para la anotación completa y automatizada de genomas de bacterias y arqueas, en el cual se realiza tanto la anotación estructural como funcional de los genes codificantes de proteínas y la identificación de ARN ribosómicos y transferentes. En el proceso, los genes se clasifican en distintos subsistemas funcionales y, con toda la información, se puede generar una reconstrucción metabólica inicial (Aziz y col., 2008).

Con Sma3s la anotación se lleva a cabo partiendo de un fichero FASTA que contiene las secuencias de aminoácidos de las proteínas definidas en la anotación estructural o la secuencia de nucleótidos de los genes correspondientes. Se trata de un método rápido, preciso y flexible para la anotación de un gran número de proteínas. Este programa se ejecuta de forma local y está constituido por tres módulos distintos (M1, M2 y M3). Cada módulo utiliza una búsqueda BLAST como punto de partida, la cual se realiza a través de los programas BLASTP o BLASTX dependiendo de si las secuencias a anotar se almacenan como proteínas o nucleótidos. Como base de datos de proteínas, Sma3s emplea UniProt (The UniProt

## Introducción

Consortium, 2007) en formato de texto plano (con extensión .dat), la cual se acota a la división taxonómica del genoma de estudio (bacterias, plantas, mamíferos, invertebrados, etc.) (Muñoz-Mérida y col., 2014).

El módulo 1 (M1) es el más restrictivo y utiliza la estrategia de Top-BLAST, es decir, el análisis del mejor resultado del BLAST, para detectar si la proteína analizada está presente en la base de datos o, al menos, una muy similar, de forma que se asigna la anotación de dicha proteína en la base de datos a la proteína de estudio. Los parámetros que exige el M1 son un 90 % de identidad y un 90 % de cobertura en el alineamiento de la secuencia de la base de datos. El módulo 2 (M2) se basa en la búsqueda de ortólogos a través de BLAST recíproco y es menos restrictivo que el M1, con un requerimiento del 75 % de identidad y una cobertura de, al menos, el 75 % de la proteína de la base de datos. Por último, el módulo 3 (M3) es el más permisivo de los tres módulos y toma la información para la anotación de múltiples secuencias homólogas utilizando una aproximación que consiste en varios pasos: primero se escanea el resultado del BLAST para buscar alineamientos estadísticamente significativos, después se agrupan en familias, se realiza una anotación preliminar combinando anotaciones individuales no redundantes de cada agrupación y, finalmente, se realiza un enriquecimiento de términos biológicos para seleccionar las anotaciones sobrerrepresentadas en cada grupo de secuencias homólogas (Muñoz-Mérida y col., 2014).

Dentro de UniProt, Sma3s da prioridad a las anotaciones procedentes de Swiss-Prot (revisadas) frente a las de TrEMBL (no revisadas). Además, el procedimiento estándar de anotación con esta herramienta es un proceso secuencial en el que las proteínas que se puedan anotar con el M1 no pasarán al M2 y, del mismo modo, sólo se anotarán con el M3 aquellas secuencias que no se hayan podido anotar con los dos módulos anteriores. Aunque este es el flujo de trabajo habitual, la flexibilidad de Sma3s permite ejecutar los módulos por separado para que el usuario pueda establecer su propio protocolo de anotación. Cabe destacar que la anotación a través de Sma3s no incluye a los ARNs funcionales no codificantes por lo que, para completar la anotación, habría que utilizar programas adicionales.

Además de los servicios y herramientas disponibles para la anotación de genomas procariotas, el *National Center for Biotechnology Information* (NCBI) ha desarrollado un procedimiento interno para la anotación completa de los genomas de bacterias y arqueas que se suben a la base de datos GenBank, si el usuario lo solicita ([https://www.ncbi.nlm.nih.gov/genome/annotation\\_prok/](https://www.ncbi.nlm.nih.gov/genome/annotation_prok/)). En la base de datos RefSeq, todos los genomas se anotan con este procedimiento. Se trata de una base de datos diversa taxonómicamente y no redundante, en la que se alberga la anotación de ADN, ARN y proteínas, incluyendo genomas completos.

Generalmente, las proteínas anotadas funcionalmente se almacenan en bases de datos, de forma que la información quede a disposición de la comunidad científica. Una de las principales y más completas bases de datos de proteínas es UniProt (<http://www.uniprot.org>), originada a partir de la fusión de los bancos de datos Swiss-Prot (*Swiss Institute of Bioinformatics*), TrEMBL (*Translation of EMBL Nucleotide Sequence Database*) y PIR (*Protein Information Resource*). UniProt se divide en dos secciones: Swiss-Prot y TrEMBL. La primera contiene anotaciones curadas y revisadas por expertos, en algunos casos con evidencias experimentales, mientras que la segunda se compone de anotaciones automáticas sin revisar (The UniProt Consortium, 2007). Cada ficha de UniProt, en formato de texto plano, contiene diversos campos de información para una misma proteína, los cuales se organizan por líneas que comienzan con un código descriptivo de dos letras mayúsculas.

En el proceso de anotación llevado a cabo por Sma3s, la herramienta toma información del campo DR de la ficha de UniProt para incorporar los términos GO (*Gene Ontology*) a la nueva proteína anotada (Ashburner y col., 2000). Estos términos proporcionan un vocabulario controlado para describir a una proteína en función de tres ontologías biológicas: función molecular, proceso biológico y componente celular. Sma3s también toma los términos del campo DR obtenidos de la base de datos InterPro (Hunter y col., 2009), los cuales aportan información sobre la familia a la que pertenece la proteína y sobre los dominios y sitios funcionales que posee. De los campos KW y CC, la herramienta de anotación adquiere, respectivamente, las palabras clave e información sobre las rutas metabólicas en las que está implicada la proteína (Muñoz-Mérida y col., 2014). En las fichas web de cada proteína almacenada en UniProt hay enlaces directos a las distintas bases de datos, que aportan nueva información sobre la proteína. En 2015 se incorporó a UniProt una página de proteomas que permite al usuario acceder al total de proteínas de un organismo secuenciado completamente (The UniProt Consortium, 2017).

Otras bases de datos de proteínas comúnmente utilizadas son Pfam (Finn y col., 2010) y PDB (*Protein Data Bank*; Bernstein y col., 1977). La primera es una base de datos de familias de proteínas y dominios que construye alineamientos múltiples de dominios proteicos comunes y proporciona información sobre la función de la proteína. La base de datos PDB es el único repositorio internacional para el procesamiento y la distribución de estructuras proteicas. Las estructuras almacenadas en esta base de datos se determinaron por cristalografía de rayos X (~86 %) o resonancia magnética nuclear (~14 %) (Xu y Xu, 2004).

Además de las bases de datos citadas, existen muchas otras y, tal y como se ha podido comprobar, cada una de ellas especializada en determinadas propiedades de las proteínas, como la función, la estructura, la clasificación en familias funcionales (basada en la secuencia,

la estructura o la función) o incluso la interacción con otras proteínas o ligandos (revisado en Xu y Xu, 2004).

### **4.3.3. Detección y anotación funcional de ARNs no codificantes**

Además de los genes codificantes para proteínas, en los sistemas vivos se transcriben genes que dan lugar a ARNs funcionales, los cuales no se traducen a proteínas sino que ejercen su función como moléculas de ARN. Entre estos ARNs se encuentran tanto los ribosómicos (ARNr) y los transferentes (ARNt), que juegan un papel fundamental en la traducción de los ARNs mensajeros (ARNm) a proteínas, como otros ARNs no codificantes con distintas funciones. Puesto que estos genes no se traducen, sino que, salvo excepciones, sólo se transcriben, se han desarrollado herramientas bioinformáticas para identificarlos que se basan en metodologías distintas a las empleadas en la búsqueda de pautas abiertas de lectura. Algunas de estas herramientas son RNAmmer (Lagesen y col., 2007) para la detección de ARNr; ARAGORN (Laslett y Canback, 2004), tRNAscan-SE (Lowe y Eddy, 1997) y tRNAfinder (Kinouchi y Kurokawa, 2006) para la anotación de ARNt, e Infernal (Nawrocki y col., 2009; Nawrocki, 2014) para la búsqueda de ARNs no codificantes en general. Con estas herramientas, se realiza a la vez tanto la anotación estructural como funcional de los genes que los codifican.

La predicción computacional de los ARN ribosómicos y transferentes es más eficaz que la del resto de ARNs funcionales no codificantes, salvo aquellos que se encuentran muy conservados. En el caso de los pequeños ARNs reguladores bacterianos (sRNA, acrónimo en inglés), esto se debe a que no presentan estructuras secundarias comunes tan evidentes como la estructura en forma de hoja de trébol característica de los ARNt o las estructuras típicas de los ARNr (Petrov y col., 2014). Además, los sRNA no contienen ninguna secuencia estadísticamente distinguible que ayude a identificarlos de forma general y no se encuentran conservados entre especies filogenéticamente distantes (Sridhar y Gunasekaran, 2013). Aún con estas dificultades, se han desarrollado métodos computacionales para la predicción de sRNA en procariotas. Algunas de estas herramientas se basan en genómica comparada, como QRNA (Rivas y Eddy, 2001), la cual emplea el análisis de estructuras consenso y la genómica comparada para encontrar sRNA en genomas bacterianos, dando peso a la conservación de las regiones intergénicas entre especies relacionadas como indicador de la presencia de posibles pequeños ARNs.

Otro programa de predicción de ARNs no codificantes basado en la genómica comparada es Infernal, el cual utiliza los modelos de covarianza (CMs, acrónimo en inglés) almacenados en la base de datos Rfam para realizar la detección de ARNs no codificantes a lo largo del genoma de estudio. Los CMs son modelos probabilísticos de la secuencia y la estructura

secundaria de una familia de ARN, los cuales se construyen a partir de alineamientos múltiples de homólogos conocidos de la familia. Infernal está constituido por los programas *cmbuild*, que se utiliza para construir un CM a partir de un alineamiento, *cmsearch*, para buscar una secuencia que se ajuste a un CM en una base de datos, como por ejemplo un genoma completo, y *cmalign*, para crear alineamientos múltiples de posibles homólogos a partir de un CM. En definitiva, para anotar ARNs funcionales con Infernal se necesitan los CMs de la base de datos Rfam, que vienen pre-calibrados, el programa *cmsearch* y el genoma completo del organismo que se quiere anotar, el cual se utiliza como base de datos para buscar secuencias que se ajusten a cada uno de los CMs analizados (Nawrocki, 2014).

El principal problema de la anotación computacional de ARNs funcionales es que es un procedimiento que no permite detectar ARNs *de novo*, es decir, cuyas secuencias y estructuras no estén almacenadas en las bases de datos. Esto no ocurre con las pautas abiertas de lectura de las proteínas, las cuales pueden predecirse estructuralmente aunque no estén almacenadas en las bases de datos de proteínas. Por tanto, muchos grupos de investigación optan por abordajes experimentales para la identificación de pequeños ARNs en genomas procariotas, como la secuenciación diferencial de ARN (dRNA-seq), que se ha utilizado para definir los inicios de transcripción (TSS, acrónimo en inglés) a lo largo del genoma y detectar nuevos pequeños ARNs en la bacteria *Helicobacter pylori* (Sharma y col., 2010; Bischler y col., 2015), la rizobacteria *Bacillus amyloliquefaciens* FZB42 (Fan y col., 2015) o la haloarquea *Haloferax volcanii* (Babski y col., 2016).

Tal y como ocurre con las proteínas, existen bases de datos de ARNs funcionales disponibles para la comunidad científica, entre las cuales encontramos bases de datos especializadas en un tipo de ARN en concreto. Existen bases de datos específicas de ARNs transferentes, como GtRNadb (Chan y Lowe, 2009) o tRNadb (Frank Jühling y col., 2009), de ARNs ribosómicos, como RDP (*Ribosomal Database Project*) (Cole y col., 2014) o SILVA (Quast y col., 2013) y de ARNs reguladores bacterianos, como el repositorio BSRD (*Bacterial Small Regulatory RNA Database*) (Li y col., 2013). Algunas bases de datos contienen información de todos los tipos de ARNs no codificantes. Tal es el caso de Rfam, en la cual se almacena una colección de familias de ARNs que incluyen ARNs ribosómicos, transferentes y otros ARNs no codificantes. Cada familia se representa por un modelo de covarianza (CM), alineamientos múltiples y estructuras secundarias consenso (Gardner y col., 2011) y, como se mencionó anteriormente, se utiliza a través de la herramienta Infernal para detectar sRNA en genomas completos.

### 5. Pequeños ARNs reguladores no codificantes en bacterias

Actualmente se sabe que la fracción transcrita de un genoma bacteriano es mucho más amplia de lo que se pensaba anteriormente y que las moléculas de ARN no solo están implicadas en el proceso de traducción de la expresión génica a proteínas sino que juegan un papel fundamental en muchos otros procesos (Wade y Grainger, 2014).

Los pequeños ARNs reguladores no codificantes constituyen una clase emergente de reguladores de la expresión génica que se encuentran distribuidos en todos los dominios de los seres vivos. En bacterias, los sRNAs reguladores controlan la expresión de genes implicados en multitud de procesos biológicos a través de distintos mecanismos de acción (Wagner y Romby, 2015).

En general, presentan un tamaño pequeño, de entre 50 y 300 nucleótidos (Storz y col., 2011) y pueden sintetizarse a partir de genes independientes, desde promotores internos en regiones codificantes, como ocurre con MicL (Guo y col., 2014), desde regiones 3' no traducidas, como por ejemplo DapZ (Chao y col., 2012), mediante el procesamiento de ARNs de mayor tamaño, como SdsN (Hao y col., 2016), desde la cadena opuesta al gen que regulan, como ocurre con los ARN antisentido, e incluso pertenecer al propio ARNm diana, como es el caso de los *riboswitches* (Tucker y Breaker, 2005). Estos últimos elementos no son pequeños ARNs *per se*, sino que forman parte de la región 5'UTR del ARNm que regulan y, como consecuencia, actúan en *cis*. Los *riboswitches* pueden adoptar conformaciones diferentes en respuesta a señales y cambios en el ambiente mediante la unión a pequeños metabolitos, con lo que pueden tanto inhibir como promover la traducción del correspondiente ARNm (revisado en Garst y col., 2011).

#### 5.1. Tipos de ARNs reguladores y mecanismos de acción

El principal mecanismo de acción de los pequeños ARNs es la regulación post-transcripcional de la expresión génica mediante la unión a los ARNm diana, lo cual puede derivar en la inhibición o la activación de su traducción. Además de este modo de acción, estos elementos pueden actuar uniéndose a proteínas reguladoras evitando que estas ejerzan su función. En este caso, el pequeño ARN contiene secuencias que mimetizan el sitio de unión de la proteína reguladora al ARN, de forma que es capaz de secuestrar al regulador y evitar que este se una a la secuencia de ARN donde ejercería su acción reguladora (Faucher y Shuman, 2011). Algunos sRNAs presentan incluso una función dual, actuando como pequeños ARNs reguladores y codificando para pequeñas proteínas (Vanderpool y col., 2011). Por último, los ARNs reguladores denominados CRISPR (del inglés *Clustered Regularly Interspaced Short Palindromic Repeats*) se constituyen de pequeñas regiones

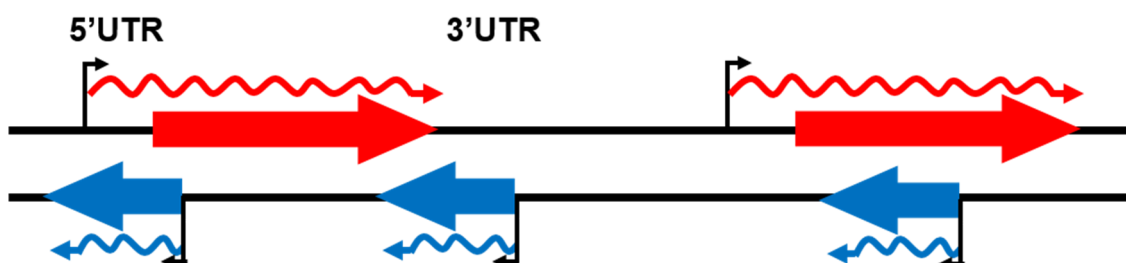


repetidas con homología a secuencias de bacteriófagos y plásmidos. Los CRISPR proporcionan a la bacteria un mecanismo de resistencia a bacteriófagos y evitan la conjugación de plásmidos (Waters y Storz, 2009).

Los sRNAs que ejercen su función mediante la hibridación con el ARNm diana pueden clasificarse en *cis* o *trans* según su relación con la diana a la que regulan. El efecto sobre el ARN mensajero puede ser tanto positivo como negativo, de forma que la unión del sRNAs cerca o en la propia secuencia de unión al ribosoma (RBS) evitaría el reconocimiento del ribosoma y la subsecuente traducción, mientras que, de forma opuesta, la unión del sRNAs puede generar un cambio en la estructura secundaria del ARNm y liberar el RBS para permitir el inicio de la traducción (Faucher y Shuman, 2011). Además, la unión de los pequeños ARNs puede inducir la degradación del ARNm a través del reclutamiento de RNAsas (Waters y Storz, 2009).

### 5.1.1. ARNs reguladores codificados en *cis*

Los sRNAs codificados en *cis* (o ARNs antisentido) se transcriben desde la cadena de ADN opuesta al ARNm diana. Además, muestran una complementariedad de bases perfecta con el ARNm, que puede llegar a ser bastante extensa (de más de 75 nt). La complementariedad entre el ARNs y su ARNm diana puede ocurrir tanto en la región codificante como en las regiones 5' y 3' no traducidas (Figura 7).



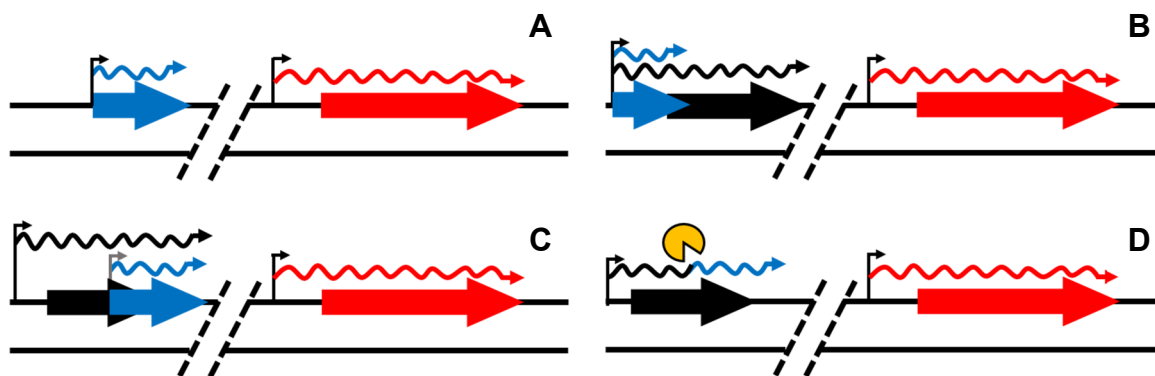
**Figura 7. Representación de pequeños ARNs codificados en *cis*.** Las líneas negras representan las cadenas de ADN y las flechas negras corresponden con los inicios de la transcripción. Los sRNAs (flecha azul) pueden ser complementarios a las regiones 5' o 3' no transcritas (UTR) (esquema de la izquierda) o a la región codificante del gen diana (flecha roja) (esquema de la derecha). Las flechas onduladas corresponden a los transcritos del gen diana (rojo) y del pequeño ARN regulador (azul). Adaptada de Svensson y Sharma, 2016.

La mayoría de ARNs antisentido se han identificado en plásmidos, elementos transponibles y bacteriófagos, donde actúan manteniendo un número de copias adecuado (Brantl, 2007). Mientras que los ARNs antisentido codificados en plásmidos e implicados en la regulación de la replicación, conjugación o segregación, se expresan constitutivamente, los

cromosómicos se expresan solo bajo ciertas condiciones, como en la fase estacionaria (por ejemplo GadY; Opdyke y col., 2004) o cuando el hierro es abundante en el medio, como IsrR (Dühning y col., 2006). El papel fisiológico de los ARNs antisentido que se expresan desde la secuencia cromosómica del genoma de las bacterias es menos conocido. Algunos promueven la degradación y/o reprimen la traducción de ARN mensajeros que codifican para proteínas que son tóxicas a concentraciones altas, actuando como sistemas toxina-antitoxina (revisado en Fozo y col., 2008) mientras que otros modulan la expresión de genes organizados en operón (Waters y Storz, 2009).

### 5.1.2. ARNs reguladores codificados en *trans*

De forma opuesta a lo que ocurre con los sRNAs codificados en *cis*, los codificados en *trans* se localizan en regiones genómicas distintas a las de sus dianas. Además, la hibridación entre el sRNA y el ARNm diana ocurre de forma discontinua y cubriendo solo una pequeña región de ambos ARNs. A pesar de que muchos sRNAs codificados en *trans* se transcriben de forma aislada desde regiones intergénicas, algunos pueden generarse desde las regiones 5' y 3' no traducidas, comenzar su transcripción dentro de regiones codificantes e incluso obtenerse a partir del procesamiento de las regiones 3'UTR de ARNs mensajeros (Miyakoshi y col., 2015; Svensson y Sharma, 2016) (Figura 8).



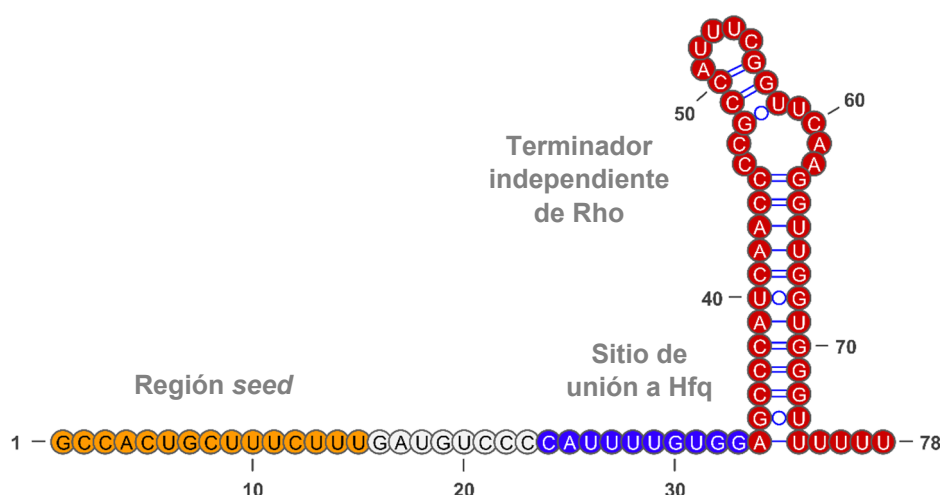
**Figura 8. Representación de pequeños ARNs codificados en *trans*.** Las líneas negras paralelas representan las cadenas de ADN y las flechas negras verticales los inicios de la transcripción. Los sRNAs (en azul) pueden sintetizarse como (A) genes independientes, (B) desde las zonas 5'UTR, (C) desde regiones codificantes y 3'UTR a partir de promotores internos (en gris) y (D) por el procesamiento del ARNm. Adaptada de Svensson y Sharma, 2016. El gen diana y su transcrito se representan en rojo y los genes de los que pueden derivar los sRNAs en negro.

Al contrario de lo que ocurre con los pequeños ARNs codificados *cis*, los codificados en *trans* suelen regular múltiples dianas. A menudo, la interacción entre el ARN regulador y el ARN mensajero ocurre en la región 5'UTR, cerca o en el sitio de unión del ribosoma, pero también puede darse en la zona codificante o en el 3'UTR (Jagodnik y col., 2017). Además,

la función de muchos sRNAs codificados en *trans* depende de la proteína de unión a ARN Hfq, la cual facilita y estabiliza la interacción entre los sRNAs y sus dianas (Waters y Storz, 2009; Updegrove y col., 2016), mientras que los pequeños ARNs codificados en *cis* no requieren, generalmente, la intervención de Hfq para llevar a cabo la unión a su diana (Brantl, 2007).

La interacción del sRNA codificado en *trans* con su diana no ocurre de forma azarosa por el simple hecho de compartir parte de la secuencia de nucleótidos. Se han descrito elementos en la estructura y la secuencia de los pequeños ARNs codificados en *trans* que aseguran la correcta hibridación con los ARNm diana. La gran mayoría de este tipo de pequeños ARN reguladores contiene una región 3' estructurada en forma de terminador de la transcripción seguida de una secuencia de múltiples uracilos (poli U), lo cual promueve una terminación de la transcripción independiente de Rho y protege al sRNA frente a 3'-exonucleasas. Un segundo dominio es el sitio de unión de Hfq, el cual suele coincidir con regiones ricas en As/Us y se encuentra cerca de la estructura terminadora. Por último, una tercera región es la que se utiliza para llevar a cabo la hibridación con el ARNm diana, la cual se denomina región semilla (*seed*, en su traducción al inglés) (Storz y col., 2011).

Curiosamente, la región *seed* de muchos pequeños sRNAs se localiza hacia su extremo 5' (Guillier y Gottesman, 2008; Papenfort y col., 2010), lo cual sugiere que esta posición podría ser importante para la función (Figura 9).



**Figura 9. Estructura general de los pequeños ARNs reguladores codificados en *trans*.** Se representan las distintas regiones de un sRNAs que se une a Hfq. Imagen adaptada de Storz y col., 2011.

Las consecuencias de una hibridación imperfecta son tanto que un sRNA puede interactuar con múltiples ARNm diana como que un ARNm puede regularse por diversos sRNAs en respuesta a distintas señales. A pesar de ello, la estrategia de hibridación por la

región *seed* hace que se mantenga la especificidad regulatoria y que un solo cambio pueda afectar a la unión (Papenfort y col., 2012).

### 5.2. Procesos biológicos regulados por pequeños ARNs en bacterias

Como ya se mencionó anteriormente, los sRNAs regulan multitud de procesos biológicos en las bacterias. Se han encontrado varios ejemplos donde este tipo de reguladores controlan la expresión de proteínas de membrana externa (OMP, del inglés *Outer Membrane Proteins*). Tanto en *E. coli* como en *Salmonella*, todas las principales porinas de membrana externa se reprimen por uno o más sRNAs (revisado en Vogel y Papenfort, 2006). Un ejemplo de esta regulación es la llevada a cabo por los pequeños ARNs MicF y MicC, los cuales reprimen la traducción de los ARNm que codifican para las porinas OmpF y OmpC, respectivamente, para ayudar recíprocamente a controlar el tamaño del poro en condiciones de cambio de osmolaridad (Andersen y col., 1989; Chen y col., 2004). También se ha estudiado en detalle la regulación de la homeostasis del hierro por parte de pequeños ARNs. En este caso, el principal ejemplo es RyhB, un pequeño ARN regulado por el represor Fur (del inglés *ferric uptake regulator*). Se ha descrito que Fur reprime la transcripción de RyhB cuando el hierro es abundante en el medio y, como consecuencia, se sintetizan las proteínas de unión a hierro que estaban reprimidas por RyhB (Massé y Gottesman, 2002).

Se han descrito algunos pequeños ARNs que regulan la síntesis de factores de transcripción. Por ejemplo, varios sRNAs regulan la expresión del factor sigma de la fase estacionaria, RpoS (revisado en Majdalani y col., 2005). Como ya se mencionó anteriormente, el control de la expresión de genes de virulencia y de proteínas tóxicas (simulando el sistema toxina-antitoxina), son procesos en los que también se ha encontrado un control mediado por pequeños ARNs reguladores.

En cuanto al transporte y metabolismo de fuentes de carbono, varios pequeños ARNs se han identificado como reguladores implicados en el fenómeno de represión catabólica. Un claro ejemplo son los pequeños ARN CrcZ y CrcY en *Pseudomonas putida*, los cuales secuestran al regulador Crc gracias a una secuencia que mimetiza su sitio de unión al ARN. La proteína Crc es un represor traduccional que reconoce una secuencia específica en algunos ARNm y evita su traducción. En condiciones de represión catabólica los niveles de CrcZ y CrcY son muy bajos y queda disponible suficiente Crc para bloquear la traducción de los genes implicados en el catabolismo de las fuentes de carbono no preferenciales. Por el contrario, en condiciones de no represión catabólica, los niveles de CrcZ y CrcY aumentan y secuestran a Crc, con lo cual se expresan los genes implicados en el consumo de la fuente de carbono no preferencial (Moreno y col., 2012). Recientemente se ha descrito que tanto la unión entre CrcZ y CrcY con Crc, como la de Crc con sus ARNm diana, se estabilizan gracias

a la intervención de la chaperona Hfq (Moreno y col., 2015). En *P. aeruginosa* se ha propuesto la misma función para CrcZ y Hfq (Sonnleithner y Bläsi, 2014).

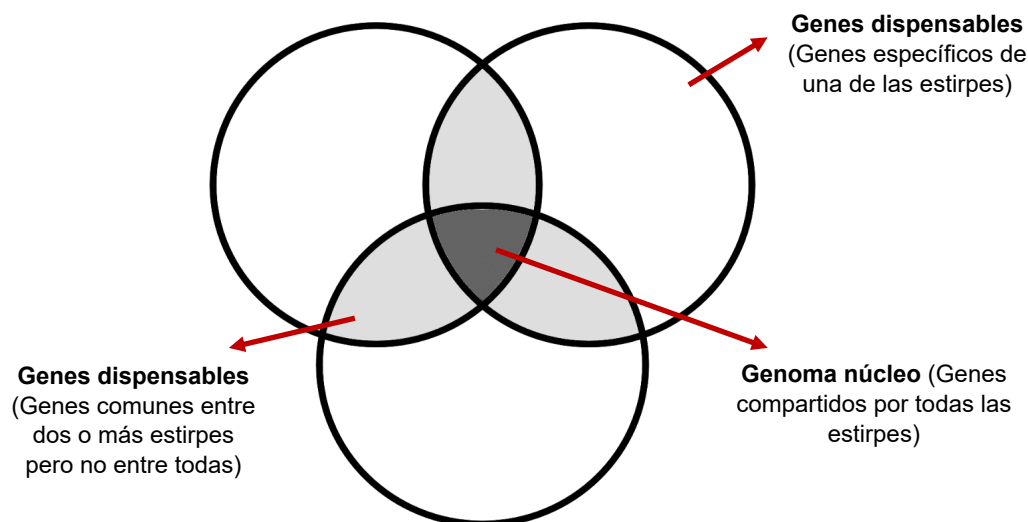
También se ha establecido en algunas bacterias, como *Yersinia pseudotuberculosis* y *E. coli*, una conexión entre el sistema regulador Csr (del inglés *Carbon Storage Regulator*) y la proteína Crp (del inglés *cAMP regulatory protein*, implicada en represión catabólica), en la cual intervienen los pequeños ARNs CsrB y CsrC (Heroven y col., 2012; Svensson y Sharma, 2016; Pannuri y col., 2016). Tanto el sistema Csr como la proteína Crp controlan el metabolismo del carbono y energético de la célula. El principal regulador del sistema Csr es la proteína CsrA, la cual se une a sus ARNm diana bloqueando la traducción o afectando a su estabilidad. Los pequeños ARNs CsrB/C contienen múltiples sitios de unión de la proteína CsrA, por lo que son capaces de secuestrar al regulador interfiriendo en su función. Por otra parte, tanto en *Yersinia* como en *E. coli* se ha descrito que, aunque a través de mecanismos distintos, la proteína Crp regula la síntesis de los ARNs CsrB/C.

## 6. Biología de sistemas

### 6.1. Genómica comparada

La disponibilidad de múltiples genomas completamente secuenciados y anotados permite seleccionar los organismos de interés y realizar comparaciones entre ellos a nivel genómico. Mientras que el análisis de un solo genoma proporciona un enorme conocimiento biológico sobre un organismo en concreto, el análisis comparativo de múltiples genomas proporciona sustancialmente más información sobre la fisiología y la evolución de las especies microbianas (Fraser y col., 2000). En definitiva, la genómica comparada es la comparación directa del material genético completo de un organismo frente al de otros, lo cual incluye la comparación del número de genes, del contenido génico, de la localización de dichos genes, de la cantidad de ADN codificante en cada genoma o el estudio de la conservación génica entre los organismos estudiados (Sivashankari y Shanmughavel, 2007).

A través de la comparación de los proteomas de distintas bacterias se puede establecer el pan-genoma y el genoma núcleo del conjunto de ellas. Normalmente, la comparación se realiza a nivel de especie, aunque hay ejemplos en los que se calcula a nivel de género (Alcaraz y col., 2010). El pan-genoma lo conforman el genoma núcleo, que contiene los genes presentes en todas las estirpes, y un genoma dispensable, que se compone tanto de genes ausentes en una o más estirpes como de genes únicos de cada estirpe (Figura 10) (Tettelin y col., 2005; Snipen y col., 2009).



**Figura 10. Representación esquemática del pan-genoma de una especie.** Se muestra el genoma núcleo en gris oscuro, los genes dispensables comunes en gris claro y los genes específicos de cada estirpe en blanco. Por simplificación, se representan solo tres estirpes, que corresponden con los tres círculos de la Figura. La imagen ha sido adaptada de Muzzi y col., 2007.

Como consecuencia, el pan-genoma reúne, de forma no redundante, el total de genes que componen la especie o el género analizado. El tamaño del pan-genoma es un indicador de la plasticidad de la especie o del género y refleja su potencial de adaptación en un medio diverso, pero hay que tener en cuenta que no se puede obtener de forma completa, puesto que es imposible secuenciar y anotar todas las estirpes que existen de la especie o el género de interés (Snipen y col., 2009).

Normalmente, los genes que se engloban dentro del genoma núcleo están asociados con el mantenimiento de los aspectos básicos de la biología de los organismos, como la replicación, traducción, síntesis de la envuelta celular o transporte (Tettelin y col., 2005). El genoma accesorio o dispensable lo conforman el conjunto de genes que son comunes entre dos o más organismos pero que no se encuentran en todos los estudiados. En este grupo se engloban genes con funciones específicas relacionadas con la supervivencia en distintos nichos y, en muchos casos, están asociados con virulencia o resistencia a antibióticos y podrían reflejar el estilo de vida dominante de los organismos analizados (Mira y col., 2010). Dentro de los genes dispensables se encuentran genes específicos de cada especie, los cuales resultan de gran interés para el estudio de nuevas funciones emergentes dentro de una especie o género. Por ejemplo, en el estudio del pan-genoma de la especie *Legionella pneumophila* se identificaron genes específicos de estirpe asociados con virulencia (D'Auria

y col., 2010). En muchos casos, los genes accesorios y los específicos de estirpe o especie proceden de eventos de transferencia horizontal (Guimarães y col., 2015).

La genómica comparada también ha revolucionado el campo de la taxonomía. A pesar de que los métodos tradicionales para la clasificación de bacterias se siguen utilizando, la disponibilidad de genomas completamente secuenciados hace posible la aplicación de nuevos abordajes comparativos para establecer relaciones filogenéticas entre microorganismos. Entre estos análisis se encuentra el alineamiento de las secuencias concatenadas de múltiples genes conservados, lo cual permite generar árboles más robustos que los obtenidos con el alineamiento de un único gen conservado, como el del ARNr 16S (Coenye y col., 2005; Martens y col., 2008). Pero no sólo pueden construirse árboles filogenéticos teniendo en cuenta genes conservados, sino también comparando los proteomas completos de las especies analizadas. Se ha definido el parámetro de media en identidad de aminoácidos o AAI (del inglés *Average Amino acid Identity*) entre dos proteomas, el cual se calcula considerando la identidad de las proteínas ortólogas entre los dos proteomas comparados y obteniendo la media. Las proteínas ortólogas se obtienen mediante la realización de BLASTP recíprocos estableciendo una identidad y cobertura de secuencia mínimos (Konstantinidis y Tiedje, 2005). Esta comparación se realiza buscando las proteínas de un organismo en el proteoma de otro, que actúa como base de datos, hasta completar todas las comparaciones posibles entre los proteomas estudiados, lo cual permite distinguir entre diferentes especies procariotas (Thompson y col., 2013).

La comparación de la secuencia nucleotídica completa de los genomas de interés también permite establecer relaciones filogenéticas a través de la obtención de la media en identidad de nucleótidos o ANI (del inglés *Average Nucleotide Identity*), de forma que se puedan distinguir a nivel de especie y de forma precisa las estirpes comparadas (Richter y Roselló-Móra, 2009). En el servidor web JSpeciesWS se puede seleccionar un conjunto de genomas y calcular el valor de ANI entre ellos (Richter y col., 2015).

## 6.2. Reconstrucción a escala genómica del metabolismo microbiano

En los últimos años, la cantidad de información a escala genómica obtenida, desde la secuenciación y anotación de genomas hasta la realización de experimentos con una aproximación -ómica (transcriptómica, proteómica, metabolómica o fluxómica), ha hecho necesario el desarrollo de herramientas bioinformáticas y análisis computacionales que permitan trabajar con la integración de todos los datos y conseguir un mejor entendimiento de los sistemas vivos. Dentro de la biología de sistemas, la reconstrucción a escala genómica del metabolismo (GEMREs, del inglés *GEnome-scale Metabolic REconstruction*), se ha

convertido en una herramienta indispensable para el estudio del metabolismo (Thiele y Palsson, 2010; Nogales, 2014).

Estas reconstrucciones son bases de datos de conocimiento específico de la especie de interés que contienen información detallada sobre el metabolismo de la misma, el conjunto de reacciones que tienen lugar en ella, la estequiometría de las reacciones, la reversibilidad, la relación entre genes, proteínas y transformaciones bioquímicas, además de otros datos bioquímicos y fisiológicos disponibles en el momento de la reconstrucción (Feist y col., 2009; Thiele y Palsson, 2010). Estas reconstrucciones pueden transformarse en modelos metabólicos a escala genómica (GEMs, del inglés *GE*nome-*s*cale *M*odels) que permiten el estudio computacional de las capacidades metabólicas y el fenotipo del organismo de interés en términos de flujos metabólicos, teniendo en cuenta una serie de asunciones (Orth y col., 2010; Nogales, 2014).

Las aplicaciones de los GEMs engloban tanto estudios de ciencia básica como aplicada, entre los que se encuentran la evolución bacteriana, la evaluación de las redes metabólicas de un sistema, la predicción de fenotipos, el descubrimiento de nuevos aspectos biológicos, como el crecimiento en diferentes fuentes de carbono, la ingeniería metabólica e incluso la interacción entre especies. En el contexto de las aplicaciones anteriores, los GEMs, junto con las herramientas de análisis disponibles, permiten generar mutantes *in silico* (adición o delección de genes) y evaluar el fenotipo de los mismos, añadir rutas heterólogas, como rutas de degradación de contaminantes o utilización de otras fuentes de carbono y determinar los efectos que tienen en el sistema, predecir genes esenciales, guiar la conexión de metabolitos huérfanos con reacciones conocidas o modelar el intercambio de metabolitos entre especies y estudiar las distintas relaciones entre ellas, como el mutualismo, la competición por los recursos del medio, el parasitismo o la evolución de la comunidad (Lewis y col., 2012; McCloskey y col., 2013).

### 6.2.1. Construcción y análisis de modelos metabólicos

Para la reconstrucción del metabolismo a escala genómica se requieren conocer las reacciones que ocurren en el organismo de interés así como los genes metabólicos implicados. El material de partida para la reconstrucción es una anotación de calidad del genoma. Además, resulta de gran utilidad la disponibilidad de datos bioquímicos y de información bibliográfica del organismo estudiado o de especies cercanas filogenéticamente.

El proceso de reconstrucción metabólica puede dividirse en 4 etapas. La primera de ellas consiste en la construcción automática de un borrador inicial del modelo a partir de la anotación del genoma, para lo cual existen distintas herramientas bioinformáticas como



Pathway Tools (Karp y col., 2002a), el servidor SEED (Henry y col., 2010) y el servidor GEMSiRV-MrBac (Liao y col., 2012). En la segunda etapa se realiza una curación manual del modelo inicial mediante la búsqueda de información en bases de datos bioquímicas y metabólicas, la realización de BLAST para encontrar similitud entre distintas proteínas y la utilización de la bibliografía disponible. La conversión de la reconstrucción a un modelo matemático se realiza en la tercera etapa del proceso y, en la etapa final, se lleva a cabo la validación y el análisis del modelo (Nogales, 2014).

De forma contraria a los modelos estocásticos y cinéticos, los cuales requieren el cálculo de una gran cantidad de parámetros y, por tanto, no son escalables a nivel genómico, los GEMs permiten obtener información metabólica detallada a través del cálculo de los flujos en una red de reacciones, aplicando una serie de restricciones (Orth y col., 2010). Por tanto, para construir y analizar estos modelos se sigue una aproximación basada en restricciones (COBRA, del inglés *CO*nstraint-*B*ased *R*econstruction and *A*nalysis), en la cual se consideran tanto los límites impuestos por el genoma del organismo estudiado (límites internos) como los debidos al medio (límites del sistema), y se describe el conjunto de fenotipos posibles bajo las condiciones especificadas (Nogales, 2014).

Un método ampliamente utilizado para analizar los GEMs es el análisis equilibrado de flujos (FBA, del inglés *Flux Balanced Analysis*). El FBA es una aproximación matemática para analizar el flujo de metabolitos a través de una red de reacciones metabólicas, con la que se puede determinar la velocidad de crecimiento de un organismo o la velocidad de producción de un metabolito de interés biotecnológico. El primer paso para aplicar el FBA es la representación matemática de las reacciones metabólicas del modelo generado. El elemento central de esta representación es una matriz numérica (denominada  $S$ ) que contiene los coeficientes estequiométricos de cada reacción. En esta matriz, cada línea representa un metabolito ( $m$ ) y cada columna una reacción ( $n$ ), por tanto, la ecuación matemática que la describe es  $S = m \times n$ .

Los coeficientes estequiométricos constituyen ya una forma de imponer restricciones en el flujo de metabolitos a través de la red de reacciones y, además, a cada una de las reacciones también se le puede establecer un límite bajo (lb, del inglés *lower bounds*) y uno alto (ub, del inglés *upper bounds*), los cuales delimitan el flujo mínimo y máximo permitidos. La matriz estequiométrica ( $S$ ) impone flujos que aseguran un balance de masa en estado estacionario, en el que la cantidad de compuestos producidos debe ser igual al total de metabolitos consumidos, lo cual establece una nueva restricción en el sistema representada por la reacción  $Sv = 0$ , donde  $v$  es el vector que representa los flujos a través de la red de

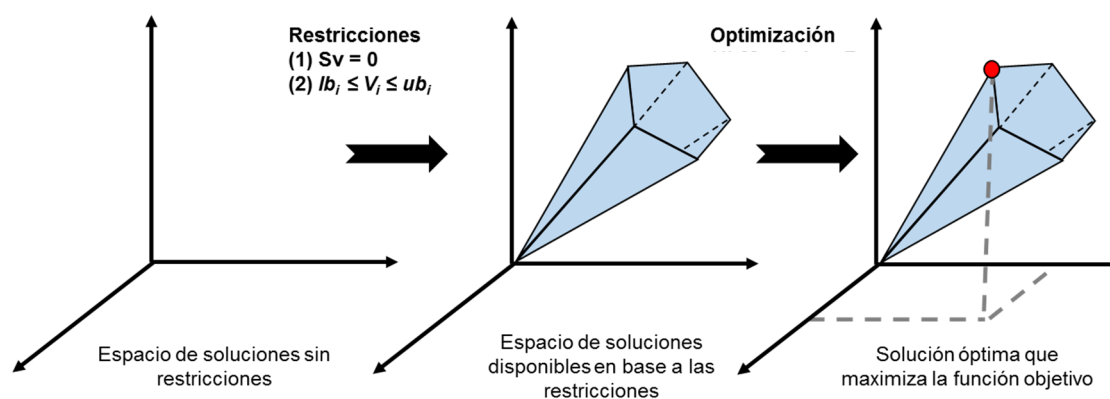
## Introducción

reacciones y tiene una longitud de  $n$ . Cada flujo puede tomar valores desde el establecido como  $lb$  hasta el indicado por  $ub$  (Orth y col., 2010; Nogales, 2014).

Tanto la matriz estequiométrica ( $S$ ) como las restricciones impuestas por los límites bajos y altos de los flujos y el establecimiento de la condición de estado estacionario definen un espacio de soluciones posibles (Figura 11). La red metabólica puede adquirir cualquier distribución de flujos que se encuentre dentro de dicho espacio.

El siguiente paso es definir un objetivo biológico que sea relevante para el problema estudiado (función objetivo). En el caso de predecir la velocidad de crecimiento, el objetivo será la producción de biomasa y, por tanto, es necesario añadir al modelo una reacción artificial denominada reacción de biomasa. En esta reacción se incluyen todos los constituyentes conocidos de la biomasa y su contribución molar al total de biomasa celular (Nogales, 2014). Esta reacción se escala para que el flujo a través de ella sea igual a la velocidad de crecimiento exponencial del organismo estudiado ( $\mu$ ).

Una vez definida la biomasa, representada en el modelo y definida como función objetivo (BOF, del inglés *Biomass Objective Function*), la aplicación del método FBA permite identificar puntos óptimos en el espacio de soluciones con los cuales se maximice o minimice la función objetivo (Figura 11) (Orth y col., 2010).



**Figura 11. Definición gráfica del concepto de modelos basados en restricciones.** Sin restricciones (izquierda), la distribución de flujos de una red metabólica puede tomar cualquier punto del espacio de soluciones. Cuando se imponen las restricciones del balance de masas establecidas por la matriz estequiométrica,  $S$ , por los límites de flujo bajo y alto aplicados a cada reacción ( $lb_i$  y  $ub_i$ , donde  $i$  representa a cada reacción) y por la asunción de estado estacionario ( $Sv = 0$ ), se define un espacio de soluciones permitido (imagen central). A través de la optimización de una función objetivo, la aplicación de FBA permite identificar una distribución de flujos óptima (derecha). Imagen adaptada de Orth y col., 2010.

El resultado de la optimización es un vector de flujos ( $v$ ) con el que se cumplen las restricciones impuestas y se maximiza (o minimiza) la función objetivo (Nogales, 2014). Existen diversos algoritmos computacionales de programación lineal, los cuales permiten obtener de forma rápida la solución óptima de grandes sistemas de ecuaciones. El paquete COBRA (Schellenberger y col., 2011) es un paquete gratuito para el entorno de programación MatLab (The MathWorks Inc., <http://www.mathworks.es/>) que se compone de distintos programas para llevar a cabo la conversión matemática y el análisis de las reconstrucciones metabólicas basadas en restricciones. Entre las herramientas de análisis de COBRA existen funciones para llevar a cabo un FBA, como la función '*optimizeCbModel*' (Orth y col., 2010), además de otras que permiten realizar el análisis de genes esenciales, generar mutantes *in silico* o llevar a cabo dFBA (análisis de flujo dinámico a lo largo de la curva de crecimiento). Para poder utilizar el paquete COBRA en MatLab se necesita que el modelo esté en formato SBML (del inglés *Systems Biology Markup Language*) (Hucka y col., 2003).

Además, los GEMs constituyen un soporte para integrar datos -ómicos que aporten información sobre la regulación de la red metabólica en una condición dada, los cuales suponen nuevas restricciones al espacio de soluciones. En un GEM inicial se asume que todas las reacciones están activas en cualquier condición. Sin embargo, muchos de los genes codificados en el genoma de la especie de interés sólo se expresan en determinadas condiciones. La incorporación de datos de expresión supone un importante refinamiento del modelo metabólico y existen protocolos diseñados para tal fin (Nogales y Agudo, 2015).

En el proceso de reconstrucción de un modelo metabólico, y en concreto en la segunda etapa de curación manual, se hace uso, además de la anotación del genoma, de la información metabólica y bioquímica almacenada en distintas bases de datos, las cuales pueden ser específicas de especies o de géneros bacterianos, como EcoCyc (Keseler y col., 2017) que recoge información sobre *Escherichia coli* K-12 o la base de datos de *Pseudomonas* (<http://www.pseudomonas.com/>) donde se reúne información de este género bacteriano (Winsor y col., 2011). Otras bases de datos almacenan información metabólica y bioquímica global de múltiples organismos. Dentro de estas últimas encontramos MetaCyc (Karp y col., 2002b), donde se describen multitud de rutas metabólicas, enzimas y las reacciones que catalizan junto con sus sustratos y productos, incorporando además información de distintas fuentes bibliográficas, las bases de datos bioquímicas KEGG (Kanehisa y Goto, 2000) y BRENDA (Schomburg y col., 2002) y la base de datos BIGG (Schellenberger y col., 2010), la cual almacena reconstrucciones metabólicas a escala genómica de distintos organismos.



## **OBJETIVOS**

---



1. Secuenciación, ensamblaje y anotación del genoma de *Sphingopyxis granuli* estirpe TFA.
2. Análisis genómico comparativo para la adscripción filogenética de *Sphingopyxis granuli* estirpe TFA y estudio de características comunes dentro del género *Sphingopyxis*.
3. Reconstrucción y análisis del metabolismo de *Sphingopyxis granuli* estirpe TFA mediante la generación de un modelo metabólico a escala genómica.
4. Identificación de elementos reguladores implicados en la represión catabólica de los genes de degradación de tetralina (*thn*) y caracterización de su mecanismo de acción.





## **MATERIALES Y MÉTODOS**



## 1. Estirpes, plásmidos, oligonucleótidos y condiciones de cultivo

### 1.1. Estirpes y plásmidos

En las siguientes Tablas se describen las estirpes bacterianas y los plásmidos empleados en el desarrollo de esta tesis.

Tabla 1. Estirpes bacterianas utilizadas durante la realización de este trabajo

Estirpe	Descripción	Referencia
<i>Escherichia coli</i>		
DH5α	F <sup>-</sup> Φ80/ <i>lacZ</i> ΔM15 Δ <i>lacZYA-argF</i> U169 <i>recA1 endA1 hsdR17</i> (r <sub>K</sub> <sup>-</sup> m <sub>K</sub> <sup>-</sup> ) <i>supE44 thi-1 gyrA relA1</i>	Hanahan, 1983
TOP10F'	F' {/ <i>lacIq</i> , Tn10(TetR)} <i>mcrA</i> Δ( <i>mrr hsdRMS mcrBC</i> ) Φ80/ <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 araD139</i> Δ( <i>ara leu</i> )7697 <i>galU galK rpsL</i> (Str <sup>R</sup> ) <i>endA1 nupG</i>	Invitrogen
<i>Sphingopyxis granuli</i> TFA		
TFA	Estirpe silvestre. Str <sup>r</sup>	Hernaez y col., 1999
TFA-1002	Fusión traduccional <i>thnC::lacZ</i> en TFA. Str <sup>r</sup> Ap <sup>r</sup>	Martinez-Perez y col., 2004
MPO209	Inserción miniTn5Km en PHA sintetasa. Str <sup>r</sup> Km <sup>r</sup>	Martin-Cabello y col., 2011
MPO209-1002	Derivado de MPO209 con integración de la fusión traduccional <i>thnC::lacZ</i> en la zona intergénica <i>thnB-thnC</i> . Str <sup>r</sup> Km <sup>r</sup> Ap <sup>r</sup>	Martin-Cabello y col., 2011
MPO501	Sustitución del gen <i>hfq</i> por <i>hfq::3xFlag</i> . Str <sup>r</sup> Km <sup>r</sup>	Esta tesis
MPO641	Fusión transcripcional al promotor de <i>suhB</i> ( <i>PsuhB::lacZ</i> ) en TFA. Str <sup>r</sup> Ap <sup>r</sup>	Esta tesis
MPO642	Sustitución del gen <i>suhB</i> por un gen de resistencia a kanamicina desde dos base aguas arriba del +1 hasta 85 bases aguas abajo del último nucleótido. Str <sup>r</sup> Km <sup>r</sup>	Esta tesis
MPO643	Derivado de MPO642 con integración de la fusión traduccional <i>thnC::lacZ</i> en la zona intergénica <i>thnB-thnC</i> . Str <sup>r</sup> Km <sup>r</sup> Ap <sup>r</sup>	Esta tesis
MPO644	Derivado de MPO209 con integración de la fusión transcripcional <i>PsuhB::lacZ</i> . Str <sup>r</sup> Km <sup>r</sup> Ap <sup>r</sup>	Esta tesis

Tabla 2. Plásmidos utilizados durante la realización de este trabajo

Plásmido	Descripción	Referencia/Fuente
pBluescript II SK (+)	Vector de clonación, Ap <sup>r</sup>	Stratagene
pEX18Tc	Vector suicida empleado para facilitar dobles eventos de recombinación, Mob <sup>+</sup> , <i>sacB</i> <sup>+</sup> , Tc <sup>r</sup>	Hoang y col., 1998
pGG1	Basado en pZE12-luc y conteniendo el epítipo 3xFLAG. Usado para introducir las regiones aguas arriba y aguas abajo de un gen de interés para marcarlo con FLAG. Km <sup>r</sup> Ap <sup>r</sup>	Laboratorio de Cynthia Sharma

Plásmido	Descripción	Referencia/Fuente
pIC552	Vector para fusiones transcripcionales con el gen híbrido <i>gal'K-lacZ</i> . Basado en pBR322. Ap <sup>r</sup>	Macián y col., 1994
pIZ1002	Fusión traduccional <i>thnC::lacZ</i> en pJES379. Contiene la región intergénica <i>thnB-thnC</i> , Ap <sup>r</sup>	Martinez-Perez y col., 2004
pIZ1016	Vector de expresión de amplio espectro conteniendo <i>lacI<sup>q</sup></i> y Ptac. Gm <sup>r</sup>	Martinez-Perez y col., 2004
pJV300	Plásmido control que expresa un ARN de ~50 nucleótidos de la cadena complementaria de la región terminadora del gen <i>rrnB</i> de <i>E. coli</i> . Bajo el promotor P <sub>LlacO1</sub> , Ap <sup>r</sup>	Sittka y col., 2007
pLAFR3	Vector de amplio espectro. Mob <sup>+</sup> , Tra <sup>-</sup> , cósmido, Tc <sup>r</sup>	Staskawicz y col., 1987
pMPO1149	Regiones aguas arriba y aguas abajo de <i>suhB</i> flanqueando a un gen de resistencia a kanamicina clonadas en pBluescript II SK (+), Ap <sup>r</sup> , Km <sup>r</sup>	Esta tesis
pMPO1150	Fragmento BamHI de 4,28 kb del vector pMPO1149 clonado en pEX18Tc, Tc <sup>r</sup> , Km <sup>r</sup>	Esta tesis
pMPO1151	Fusión transcripcional al promotor de <i>suhB</i> <i>PsuhB::lacZ</i> en pIC552, Ap <sup>r</sup>	Esta tesis
pMPO1152	pLAFR3 conteniendo el gen <i>suhB</i> junto con su región promotora, Tc <sup>r</sup>	Esta tesis
pMPO1153	pZE12-luc con <i>suhB</i> bajo el control de P <sub>LlacO1</sub> , Ap <sup>r</sup>	Esta tesis
pMPO1154	pXG10-SF con 5'UTR <i>thnR</i> bajo el control de P <sub>LtetO1</sub> , Cm <sup>r</sup>	Esta tesis
pMPO1155	pUC19 con <i>suhB</i> bajo el control del promotor de T7 para transcripción <i>in vitro</i> , Ap <sup>r</sup>	Esta tesis
pMPO1156	pUC19 con 5'UTR- <i>thnR</i> bajo el control del promotor T7 para la transcripción <i>in vitro</i> , Ap <sup>r</sup>	Esta tesis
pMPO1157	pUC19 con la región 5'UTR- <i>thnR</i> modificada bajo el control del promotor T7 para la transcripción <i>in vitro</i> , Ap <sup>r</sup>	Esta tesis
pMPO1806	Derivado de pGG1 que contiene el gen <i>hfq</i> de TFA fusionado al epítipo 3xFLAG expresado desde su propio promotor. Km <sup>r</sup> Ap <sup>r</sup>	Esta tesis
pMPO1808	Derivado de pMPO1806 que contiene la región aguas abajo de <i>hfq</i> de TFA, clonada antes del inicio del gen de resistencia a kanamicina. Km <sup>r</sup> Ap <sup>r</sup>	Esta tesis
pMPO1819	Derivado del pIZ1016 que contiene 564 pb del gen <i>hfq</i> de TFA fusionado al epítipo 3xFLAG bajo el control de P <sub>tac</sub> . Gm <sup>r</sup>	Marina Rebollo (no publicado)
pRK2013	Plásmido auxiliar para conjugaciones, Tra <sup>+</sup> , Km <sup>r</sup>	Figurski y Helinski, 1979
pUC19	Vector de clonación, Ap <sup>r</sup>	Yanisch-Perron y col., 1985
pXG0	Vector control para realizar fusiones traduccionales a GFP. P <sub>LtetO1</sub> , Cm <sup>r</sup>	Corcoran y col., 2012
pXG10-SF	Vector para fusiones traduccionales a una variante de GFP <i>superfolder</i> . P <sub>LtetO1</sub> , Cm <sup>r</sup>	Corcoran y col., 2012
pZE12-luc	Vector de expresión bajo el promotor P <sub>LlacO1</sub> , Ap <sup>r</sup>	Lutz y Bujard, 1997

## 1.2. Oligonucleótidos

A continuación se describen los oligonucleótidos empleados en el desarrollo de esta tesis doctoral, a excepción de los usados para el ensamblaje del genoma de TFA, los cuales se listan en el Anexo I.

**Tabla 3. Oligonucleótidos utilizados durante la realización de este trabajo (salvo los empleados en el proceso de ensamblaje del genoma de TFA)**

Oligonucleótido	Secuencia 5'-3'	Diana
<b>5thnR_mut_fwd</b>	CCCGGAATTCTAATACGACTCACTATAGGTACAGA CCTGATGTGCTCGCGACTGGGGCCCCAACGGTGAT TGAGGTTTCATTTTCGTAAGTAAGCCCTTAAATT GTGGG	Desde el +1 de <i>thnR</i> , contiene el promotor T7 y la zona de interacción con SuhB modificada
<b>5thnR_mut_rev</b>	CCCAAGCTTTTAAATTGAGATCCAGTTTCAGGAGGT CAATCAATTCCCACAATTTAAAGGGCTTACTTACG AAAATGAAACCTCAATCACCGTTTGGGCCCCAGTC GCGAG	Gen <i>thnR</i> , contiene zona de interacción con SuhB modificada
<b>CSO-0073</b>	CTAACAAGCTTTTCATCTACGCA	En orientación reversa. Para amplificar pGG1
<b>CSO-0075</b>	TCCTTCACAAAGAAGGGG	Entre <i>aphA3</i> y sitio XbaI. En el sentido de la secuencia para amplificar pMPO1806
<b>f27</b>	AGAGTTTGATCMTGGCTCAG	Gen codificante ARNr 16S
<b>GTPasedown_Xba</b>	GTTTTTCTAGAGGTTGACTTTCCAGCATTCG	Aguas abajo de <i>hfq</i>
<b>Hfq_end_NcoI_TFA</b>	GGCGACTGATCCCATGGACGACGC	Aguas abajo de <i>hfq</i>
<b>JVO5142</b>	GACTACAAAGACCATGACGG	En el sentido de 3xFLAG. Para amplificar pGG1
<b>KmFw_pk18</b>	GATTGAACAAGATGGATTGC	Interno a región codificante de <i>kan</i>
<b>KmRev_pk18</b>	CGTCAAGAAGGCGATAGAAGG	Interno a región codificante de <i>kan</i>
<b>PllacoB_1</b>	CGCACTGACCGAATTCATTAAAG	Gen <i>luc</i> de pZE12-luc
<b>PllacoD</b>	GTGCTCAGTATCTTGTATCCG	Reverso al promotor P <sub>LlacO1</sub> , del vector pZE12-luc
<b>pMPO1806_NcoI</b>	TATTCCTCCATGGTACTAAAACA	Al final de <i>aphA3</i> . En sentido reverso para amplificar pMPO1806
<b>pXG10_ups_gfp_F</b>	GATATCGACGTCTAAGAAAC	Aguas arriba del gen <i>sfGFP</i> en el vector pXG10-SF
<b>r519</b>	GTATTACCGCGGCTGCTG	Gen codificante ARNr 16S
<b>sfGFP_pXG10_out_R</b>	GTGCCCATTAACATCACCATC	Gen <i>sfGFP</i> en el vector pXG10-SF
<b>suhB_compF</b>	ATATAAGCTTGGGCAATTCGGCACGGACG	Tras la última base de <i>suhB</i>
<b>suhB_compR</b>	ATATAGGATCCCCATCGTCCCAATCTAGGC	Aguas arriba del promotor de <i>suhB</i>
<b>suhB_F_phosp</b>	GCCTTTTAGGCATCCTCTCC	Desde el +1 de la transcripción del gen <i>suhB</i> . Fosforilado en el extremo 5'
<b>suhB_F1_BH</b>	ATATAGGATCCGTTTCACCCAGGTGCTGAAG	Aguas abajo de <i>suhB</i>

Oligonucleótido	Secuencia 5'-3'	Diana
suhB_F2_HD	ATATACCATGGGTGTGATATTGCACTGCACAATG	Aguas arriba de <i>suhB</i>
suhB_IVT_R	CCCAAGCTTTAAAAAAGGGCCGCCGAG	Gen <i>suhB</i>
suhB_IVT_T7_F	CCCGGAATTCTAATACGACTCACTATAGGCCTTTT AGGCATCCTCTCC	Gen <i>suhB</i> , contiene promotor T7
suhB_northern	GGGAGAGGATGCCTAAAAGG	Gen <i>suhB</i>
suhB_R_xbaI	ATATATCTAGATCAAGTCTTGGGATATATG	Aguas abajo de <i>suhB</i>
suhB_R1_HD	ATATACCATGGGCGGCATATATCCCAAAGAC	Aguas abajo de <i>suhB</i>
suhB_R2_BH	ATATAGGATCCATCAGGCACCCGACATTGG	Aguas arriba de <i>suhB</i>
TFA_5S	CGAGTTCGAGATGGGATCG	Gen del ARN 5S
TFA_Hfq_Flag	GTCGCCGGAGTCGCCTTC	Región c-terminal de <i>hfq</i> . Fosforilado en el extremo 5'
TFA_Hfq_up_AatII	GTTTTTGACGTCGACCGGTTTCGAAACGACAC	Aguas arriba de la región c-terminal de <i>hfq</i>
thnR_F_Nsil	ATATAATGCATTACAGACCTGATGTGCTCGC	Desde el +1 de la transcripción del gen <i>thnR</i>
thnR_IVT_R	CCCAAGCTTTAAATTGAGATCCAGTTTCAGGAG	Región codificante del gen <i>thnR</i>
thnR_IVT_T7_F	CCCGGAATTCTAATACGACTCACTATAGGTACAGA CCTGATGTGCTCGC	Desde el +1 de <i>thnR</i> , contiene el promotor T7
thnR_R_NheI	ATATAGCTAGCAACTTTTCTGTCCACCATC	Gen <i>thnR</i>
thnR_solap_ATGs_F	GAAAGGGCAAATGTGGGAAATGATTGACCTC	Codones TTG iniciales de <i>thnR</i> cambiados a ATG
thnR_solap_ATGs_R	GAGGTCAATCATTTCCACATTTGCCCTTTC	Complementario a thnR_solap_ATGs_F

### 1.3. Medios y condiciones de cultivo

Las distintas estirpes de *E. coli* se cultivaron en medio Luria-Bertani (LB) (Sambrook y col., 1989) tanto en cultivos líquidos como en placa, con adición de antibióticos y otros elementos según fueran necesarios. Se incubaron a 37 °C y, en el caso de los cultivos líquidos, en agitación a 180 rpm.

LB: Bactotripton 10 g/L, extracto de levadura 5 g/L, NaCl 5 g/L. En el caso de LB sólido se añade agar 15 g/L. Autoclavar.

Las estirpes de *S. granuli* TFA se cultivaron en medio rico MML (Hernández y col., 1999) o medio mínimo MM (Dorn y col., 1974) suplementado con las fuentes de carbono que se listan a continuación (Tabla 4). En el caso de crecimiento anaerobio el medio se complementó con 20 mM de nitrato como aceptor de electrones. Los cultivos de *S. granuli* TFA se incubaron a 30 °C y 180 rpm de agitación, salvo los cultivos en anaerobiosis que se incubaron de manera estática.

Tabla 4. Fuentes de carbono probadas en TFA

Compuesto	Concentración	Compuesto	Concentración
$\beta$ -hidroxibutirato ( $\beta$ -HB)	40 mM u 8 mM	Palmitato	10 mM
Ácido sebácico	16 mM	Formaldehído	10 mM
Tetralina	Medio saturado	3-fenilpropanoato	5 mM
L-Lactato	53 mM	3-(3-hidroxi-fenil)propanoato	5 mM
L-Prolina	32 mM	L-Tirosina	5 mM
Etanol	0,5 % (v/v)	L-Triptófano	5 mM
Ácido oleico	10 mM	L-Fenilalanina	5 mM
Ácido pimélico	10 mM	L-Glutamina	10 mM
Octanoato	10 mM	L-Asparagina	10 mM
Butanoato	10 mM	D-Fructosa	0,2 % (p/v)
Acetato	16 mM	D-Glucosa	0,2 % (p/v)

MML: Triptona 2 g/L, extracto de levadura 1 g/L. En el caso de LB sólido: agar 15 g/L. Autoclavar. Añadir Solución 1 al 2 % (v/v) y solución 2+3 al 2 % (v/v).

MM: H<sub>2</sub>O bidestilada (y agar 20 g/L en el caso de medio sólido). Autoclavar. Añadir Solución 1 al 2 % (v/v), solución 2+3 al 2 % (v/v) y fuente de carbono.

Solución 1: Na<sub>2</sub>HPO<sub>4</sub>·12H<sub>2</sub>O 3 % (p/v), K<sub>2</sub>HPO<sub>4</sub> 5 % (p/v). Autoclavar.

Solución 2: Ca(NO<sub>3</sub>)<sub>2</sub>·4H<sub>2</sub>O 0,25 % (p/v), (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 5 % (p/v), MgSO<sub>4</sub>·7H<sub>2</sub>O 1 % (p/v), hierro amonio (III) citrato 0,05 % (p/v). Autoclavar.

Solución 3: ZnSO<sub>4</sub>·7H<sub>2</sub>O 0,1 % (p/v), MnCl<sub>2</sub>·4H<sub>2</sub>O 0,03 % (p/v), H<sub>3</sub>BO<sub>3</sub> 0,3 % (p/v), CoCl<sub>2</sub>·6H<sub>2</sub>O 0,2 % (p/v), CuCl<sub>2</sub>·6H<sub>2</sub>O 0,01 % (p/v), NaMoO<sub>4</sub>·2H<sub>2</sub>O 0,03 % (p/v). Filtrar.

Solución 2+3: Mezclar 190 mL de solución 2 con 10 mL de solución 2+3.

Debido a su carácter lipofílico, la tetralina (1,2,3,4-tetrahidronaftaleno) presenta una baja solubilidad en agua por lo que se suministró en forma de vapor en unos botes de cultivo adaptados para ello (Figura 12). La tetralina se añadió a un tubo Eppendorf que presentaba cuatro perforaciones alrededor de la línea de 0,5 mL, lo cual permite la salida de los vapores y la saturación del medio. Una vez saturados los medios con tetralina, se les añadieron los antibióticos y se inocularon con la estirpe de TFA correspondiente.



**Figura 12.** Bote diseñado para el suministro de tetralina en fase gas en cultivos líquidos de *Sphingopyxis granuli*.

## Material y Métodos

Las concentraciones de antibióticos y otros compuestos que se añadieron a los distintos medios de cultivo se listan en la Tabla 5.

Tabla 5. Concentraciones de antibióticos y otros agentes químicos añadidos a los medios de cultivo

Compuesto	Concentración en el medio	
	<i>E. coli</i>	<i>S. granuli</i>
Ampicilina	100 mg/L	5 mg/L
Estreptomicina	25 mg/ L	50 mg/L
Gentamicina	10 mg/L	--
Kanamicina	25 mg/L	10 mg/L
Tetraciclina	10 mg/L	5 mg/L
Cloranfenicol	15 mg/L	--
X-gal	25 mg/L	25 mg/L
IPTG	1 mM	--
Sacarosa	--	10 % (p/v)

### 1.4. Conservación del material biológico

Para la conservación a largo plazo de las estirpes bacterianas se procedió a su congelación a -80 °C en medio rico con glicerol al 15 % (v/v). Para la conservación de los plásmidos construidos se transformaron en la estirpe *E. coli* DH5α o TOP10F' (en los casos en los que se especifica esta estirpe) y se congelaron del mismo modo.

## 2. Secuenciación del genoma de *S. granuli* TFA

### 2.1. Tecnología de secuenciación 454

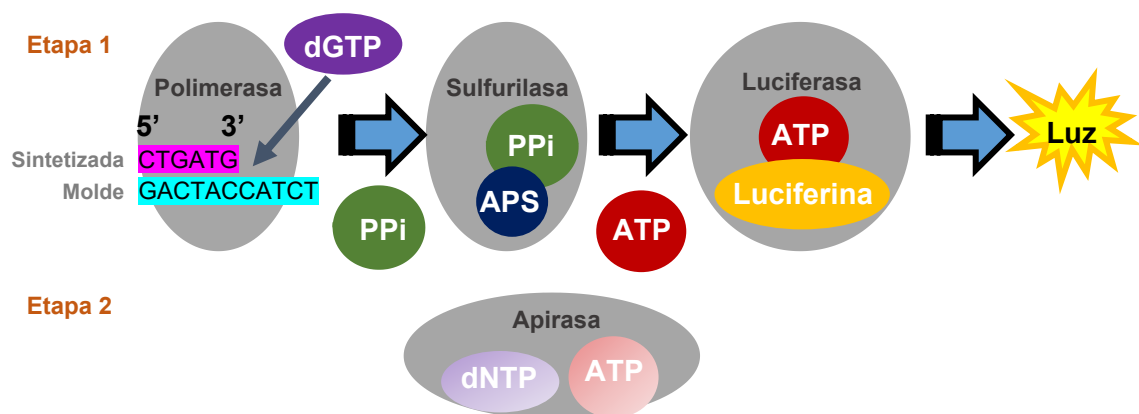
El genoma de TFA fue secuenciado en colaboración con la empresa LifeSequencing (<http://www.lifesequencing.com/>) empleando la tecnología 454. Esta plataforma realiza la secuenciación por síntesis de ADN, utilizando como molde pequeños fragmentos de cadena sencilla provenientes de la fragmentación del ADN total que se quiere secuenciar. La incorporación de un nucleótido en dicha síntesis libera un pirofosfato (PPi) que es detectado gracias a una serie de reacciones enzimáticas.

Inicialmente, a la hebra de ADN de cadena sencilla se le unen unos adaptadores comunes especializados para la técnica y de secuencia conocida, los cuales se hibridan con el cebador de la secuenciación. Uno de los adaptadores está biotinilado, lo que permite recolectar las cadenas sencillas de ADN en bolas magnéticas. Las cadenas se amplifican e inmovilizan de forma individual en gotas procedentes de una emulsión, las cuales contienen las bolas que capturan el ADN y todos los agentes necesarios para llevar a cabo la reacción de PCR. Las



bolas con el ADN inmovilizado se cargan en pocillos individuales de una superficie de fibra óptica (Rothberg y Leamon, 2008).

En la reacción de polimerización del dNTP correspondiente se libera un pirofosfato (PPi). Este pirofosfato, en presencia de adenosin 5' fosfosulfato (APS), es convertido por la enzima ATP sulfurilasa a ATP, el cual proporciona energía a la enzima luciferasa para la conversión de luciferina a oxiluciferina. Este compuesto genera luz visible de forma proporcional a la cantidad de ATP generado en la reacción anterior que, a su vez, es proporcional al número de nucleótidos incorporados. Antes de iniciar el paso de secuenciación con el siguiente nucleótido, se degradan con la enzima apirasa los dNTPs no incorporados y el exceso de ATP. El proceso se repite con los cuatro nucleótidos hasta determinar la secuencia complementaria de la cadena molde. Finalmente, la señal luminiscente es detectada por un fotodiodo o por una cámara CCD y se representa gráficamente generando un picograma (Figura 13) (Margulies y col., 2005).



**Figura 13. Esquema de la cascada enzimática que tiene lugar en la pirosecuenciación.** Se representan las enzimas implicadas (polimerasa, sulfurilasa, luciferasa y apirasa) y sus sustratos y productos.

## 2.2. Ensamblaje de las lecturas y unión de los fragmentos generados

Las pequeñas lecturas que se obtuvieron en el procedimiento de secuenciación se ensamblaron por la propia empresa LifeSequencing mediante dos tipos de ensambladores, Newbler (454 Life Sciences) y Celera Assembler (Myers y col., 2000) obteniendo el genoma de TFA en 42 fragmentos de longitud variable. Posteriormente, el orden de dichos fragmentos se predijo siguiendo las siguientes estrategias:

## **Materiales y Métodos**

- BLASTN de los extremos de los fragmentos para detectar regiones solapantes.
- BLASTX frente a la base de datos de bacterias de Uniprot de los extremos 5' y 3' de los 42 fragmentos para encontrar los extremos de proteínas truncadas.
- Comparación con las secuencias de los genomas más similares que se encontraban en las bases de datos en el momento del ensamblaje, como el de *Sphingopyxis alaskensis* RB2256, para predecir el orden de los genes en TFA.

Una vez que se obtuvo alguna evidencia de que los extremos de dos fragmentos pudieran ser contiguos, se diseñaron oligonucleótidos que hibridasen en los mismos para amplificar por PCR utilizando como molde ADN total de TFA (ver apartado 6.1.8 de materiales y métodos). Si el resultado de la PCR era positivo, se llevaba a cabo la secuenciación del fragmento (ver apartado 6.1.10 de materiales y métodos) para completar la secuencia. Para los casos en los que ninguna de las estrategias produjese un resultado positivo o no se dispusiese de hipótesis de unión entre los fragmentos, se plantearon tres nuevas aproximaciones:

- Diseño de oligonucleótidos en los extremos de los fragmentos que quedasen por unir para utilizarlos de manera combinada en reacciones de amplificación por PCR desde ADN total de TFA.
- Diseño de una sonda con la secuencia del extremo de uno de los fragmentos para, mediante hibridación con una genoteca de TFA, identificar cósmidos que portasen la región de ADN contigua a la secuencia del extremo (ver sección 6.1.7 de materiales y métodos).
- Realización de PCRs de la secuencia del extremo de un fragmento, utilizando como molde agrupaciones de cósmidos aislados de la genoteca de TFA. Las PCRs se repiten de forma sucesiva reduciendo el número de agrupaciones candidatas y, finalmente, aislando un cósmido que contenga el extremo del fragmento y la secuencia contigua.

Por último, para resolver uniones entre los fragmentos que no se pudieron ensamblar por los procedimientos descritos previamente, se mapearon en ellos las lecturas de la secuenciación mediante el software BWA (*Burrows-Wheeler Aligner*) (Li y Durbin, 2009) y el alineamiento se visualizó con IGV (*Integrative Genome Viewer*) (Thorvaldsdóttir y col., 2013). De este modo, se pretendían detectar lecturas no mapeadas que pudiesen corresponder a las regiones entre los fragmentos no unidos o mapeos anómalos que indicasen un ensamblaje

inapropiado, para finalmente cerrar el genoma. El procedimiento de ensamblaje se representa en la Figura 14.

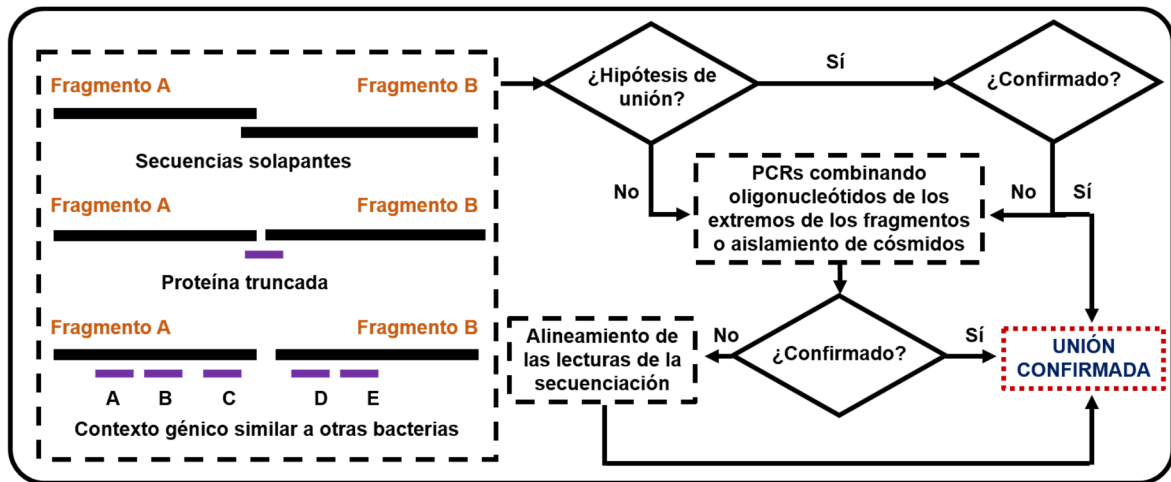


Figura 14. Esquema del flujo de trabajo seguido para la obtención de las distintas hipótesis de unión entre los fragmentos del genoma de TFA.

### 3. Herramientas bioinformáticas y bases de datos para la anotación del genoma de *S. granuli* TFA y la reconstrucción metabólica

#### 3.1. Anotación del genoma de *S. granuli* TFA

Este procedimiento se dividió en tres etapas principales: (1) la anotación estructural en la que se identificaron las pautas abiertas de lectura, (2) la anotación funcional en la que a cada proteína se le asignó una descripción completa de su función y (3) identificación y clasificación de ARNs no codificantes (Figura 15).

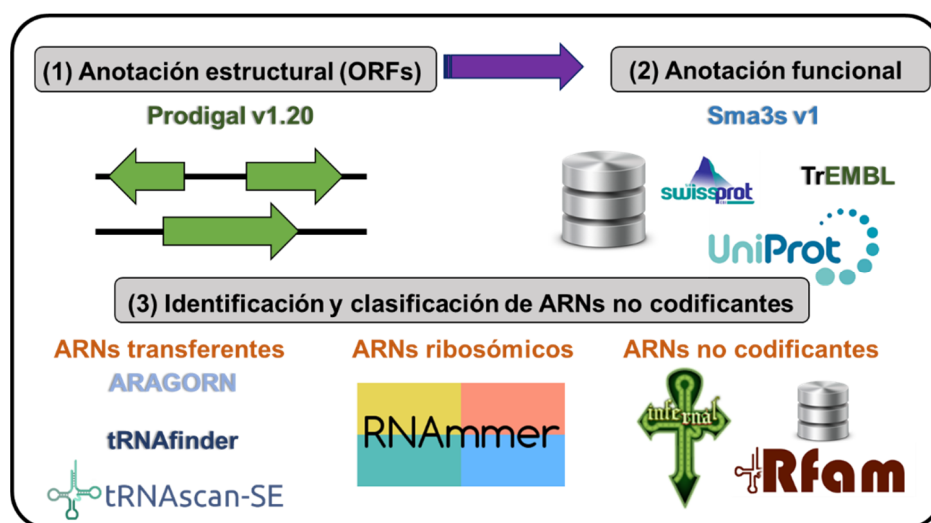


Figura 15. Esquema del proceso de anotación. Se muestran las bases de datos empleadas para la anotación de los genes codificantes para proteínas (1 y 2) y para los ARNs no codificantes.

### **3.1.1. Anotación estructural y funcional de genes codificantes**

Para la anotación estructural de proteínas se empleó el servidor Prodigal v1.20, el cual presenta una alta especificidad y precisión en la detección de pautas abiertas de lectura en genomas microbianos (Hyatt y col., 2010). Se consideraron los tres posibles codones de inicio predominantes en bacterias (ATG, GTG, and TTG). Posteriormente, las proteínas detectadas se anotaron funcionalmente con una versión modificada del programa de anotación modular Sma3s v1 (Muñoz-Mérida y col., 2014), el cual ha demostrado una alta precisión con genomas bacterianos. Sma3s utiliza la base de datos Uniprot para asignar a cada entrada el nombre de gen, la descripción y el número EC (*Enzyme Commission*). Además, añade información adicional sobre los términos GO (*Gene Ontology*), palabras clave, rutas metabólicas y motivos y dominios de InterPro.

En el caso de la anotación del genoma de *S. granuli* TFA, se siguió un protocolo de anotación modificado para obtener una alta sensibilidad y calidad dando preferencias a las anotaciones procedentes de la base de datos curada SwissProt (The UniProt Consortium, 2007). Además, el fichero FASTA que contenía las secuencias proteicas definidas por Prodigal se usó como archivo de entrada para realizar 6 ejecuciones distintas del programa Sma3s: 3 frente a la base de datos SwissProt de Bacterias (manualmente curada) y las otras 3 frente a la base de datos TrEMBL (anotada automáticamente), utilizando los tres módulos de Sma3s de forma independiente en ambos grupos:

- Módulo A1 de Sma3s: Busca a la propia proteína en la base de datos.
- Módulo A2 de Sma3s: Busca proteínas homólogas a la secuencia de entrada.
- Módulo A3 de Sma3s: Toma la información de múltiples secuencias homólogas.

Los resultados de las 6 ejecuciones se combinaron, con la ayuda de un script escrito en lenguaje Perl, para enriquecer el conjunto final de anotaciones, dando prioridad a las procedentes de SwissProt y a las procedentes de los Módulos A1, A2 y A3, por ese orden. En el proceso se evitaron las anotaciones no caracterizadas y las que presentaban nombres de genes no informativos. Tras esta anotación, se llevó a cabo una adicional con el servidor WebMGA (Wu y col., 2011) para asignar categorías COG a cada proteína. Para incrementar el número de genes anotados, se tomaron las proteínas que, tras la anotación inicial, seguían sin función asignada y se realizó una búsqueda manual de proteínas similares en la base de datos no redundante del NCBI, a través de la herramienta BLASTP. Se asignó la anotación del mejor resultado siempre y cuando se cumpliesen los siguientes parámetros: más del 30 %

de identidad, más del 70 % de la proteína cubierta en el alineamiento y un e-value  $\leq$  a  $1e-05$ . Finalmente se eliminaron de la anotación las proteínas con una longitud menor de 50 aminoácidos que estuviesen descritas como no caracterizadas.

### 3.1.2. Anotación de ARNs no codificantes

Para la detección de ARN no codificantes en el genoma de TFA se utilizaron distintas herramientas dependiendo del tipo de ARN a detectar. Para los ARN transferentes se emplearon tres programas distintos: ARAGORN (Laslett y Canback, 2004), tRNAscan-SE (Lowe y Eddy, 1997) y tRNAfinder (Kinouchi y Kurokawa, 2006). Por otra parte, la predicción de ARN ribosómicos se realizó mediante RNAmmer (Lagesen y col., 2007). Finalmente, se llevó a cabo una búsqueda global de ARNs no codificantes con el programa *cmsearch* de la herramienta Infernal v1.1 (Nawrocki y Eddy, 2013), empleando como base de datos los modelos de covarianza de Rfam 12.0 (Nawrocki y col., 2015) y descartando las predicciones con un e-value  $\leq$   $1e-05$ . En todos los casos los parámetros se mantuvieron por defecto.

Para analizar la estructura secundaria de pequeños ARNs se utilizó el servidor web RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) (Lorenz y col., 2011) y para la edición del resultado la herramienta VARNA (Darty y col., 2009). En RNAfold se introdujo la secuencia de ARN y se mantuvieron los parámetros por defecto para obtener la estructura, la cual se descargó en formato PNG. Adicionalmente, la estructura se descargó en formato VIENNA (basado en paréntesis y puntos) y se utilizó como dato de entrada en VARNA para editarla.

### 3.1.3. Identificación del origen de replicación

Para predecir el origen de replicación, OriC, se utilizó la herramienta Ori-Finder (Gao y Zhang, 2008). Esta herramienta se basa en el análisis de la asimetría en la composición de bases del genoma utilizando el método Z-curva (Gao, 2014, Zhang y Gao, 2017), la distribución de las cajas DnaA donde se une la proteína DnaA para iniciar la replicación y la presencia de genes que se encuentran frecuentemente cerca del origen de replicación, como *dnaA*, *dnaN*, *hemE* o *gidA*. En las Z-curvas se representa la disparidad entre purinas (R) y pirimidinas (Y),  $RY=(A+G)-(C+T)$ , la disparidad entre las bases con grupo amino (M) y ceto (K),  $MK=(A+C)-(G+T)$ , la disparidad entre A y T,  $AT=A-T$  y la disparidad entre G y C,  $GC=G-C$ . Todos estos parámetros se calculan de forma acumulativa desde una base determinada, base 0, hasta otra base, base N, del genoma, donde N toma valores desde 0 hasta la longitud del genoma. RY representa el exceso de purinas frente a pirimidinas, MK el exceso de bases con grupo amino frente a bases con grupo ceto, AT el exceso de A frente a T y GC el exceso

de G frente a C. Como ficheros de entrada a la herramienta se utilizaron la secuencia del genoma en formato FASTA y la anotación en formato PTT y se consideró como consenso la caja DnaA de *E. coli* permitiendo únicamente una base desigual.

### **3.1.4. Identificación de islas genómicas, profagos y secuencias CRISPR**

Para detectar islas genómicas en el genoma de TFA se utilizó el programa IslandViewer 3 (Dhillon y col., 2015). La herramienta se basa en tres métodos distintos IslandPick, IslandPath-DIMOB y SIGI-HMM, de los cuales solo IslandPath-DIMOB (Hsiao y col., 2005) y SIGI-HMM (Waack y col., 2006) se han considerado en esta tesis doctoral. El primero define las islas genómicas según la diferencia en composición de dinucleótidos y la presencia de elementos móviles mientras que el segundo considera el distinto uso de codones. El fichero de entrada fue la secuencia y anotación del genoma de TFA en formato GBK (GenBank) y se mantuvieron los parámetros por defecto.

Para la localización de profagos se empleó el programa PHAST (Zhou y col., 2011), que utiliza una base de datos para buscar proteínas relacionadas con fagos en el genoma analizado. Tras la búsqueda de estas proteínas, se determina si existen agrupaciones suficientemente cercanas de las mismas y si la región definida tiene una amplitud adecuada para constituir un profago. Como entrada se utilizó el fichero GBK con la secuencia y la anotación del genoma de TFA.

El programa web CRISPRfinder (Grissa y col., 2007) se utilizó para encontrar secuencias CRISPR (del inglés, *Clustered Regularly Interspaced Short Palindromic Repeats*) en el genoma de TFA. Esta herramienta utiliza un protocolo para encontrar los elementos característicos de las secuencias CRISPR, como las pequeñas regiones repetidas y los espaciadores. El fichero de entrada utilizado fue la secuencia genómica de TFA en formato FASTA.

## **3.2. Reconstrucción metabólica de *S. granuli* TFA**

### **3.2.1. Obtención del borrador inicial del modelo**

Para construir un modelo inicial de TFA se utilizó como referencia la información disponible en los modelos metabólicos de *Escherichia coli* K12 (K12) y *Pseudomonas putida* KT2440 (KT2440) (los modelos bacterianos más completos construidos hasta el momento). Para ello, se identificaron sistemáticamente los genes ortólogos entre TFA y las dos bacterias anteriores mediante BLASTP recíprocos. Las listas de ortólogos generadas se usaron como ficheros de entrada para el servidor GEMSiRV-MrBac (Liao y col., 2012), junto al

correspondiente modelo de referencia en formato SBML (*Systems Biology Markup Language*). Como ficheros de salida del servidor se obtuvieron dos borradores del modelo de TFA en formato SBML, uno basado en K12 y otro en KT2440, los cuales contenían las reacciones presentes en los modelos de K12 y KT2440 asignadas a los genes ortólogos previamente identificados en TFA.

Un tercer borrador del modelo se realizó de manera automática utilizando el servidor SEED (Henry y col., 2010), el cual contenía toda la información metabólica de TFA extraída de su secuencia genómica. Finalmente, los tres borradores se exportaron a Excel para manejarlos más fácilmente.

Para concluir la primera etapa del procedimiento de reconstrucción, los borradores generados se combinaron eliminando la redundancia y creando el modelo inicial de TFA.

### **3.2.2. Curación manual del modelo inicial**

Antes de continuar con el proceso de reconstrucción, se realizó una revisión manual de toda la información contenida en el modelo inicial y, además, se incluyeron capacidades metabólicas conocidas de TFA que, por su especificidad, no hubieran sido incorporadas en los borradores. Para la curación manual se establecieron las siguientes etapas:

#### **Verificación de las reacciones incluidas en el modelo inicial**

Se comprobaron una por una todas las reacciones incluidas en el borrador con especial atención en revisar las asociaciones entre gen-proteína-reacción (GPR) para corregir posibles errores, los cuales llevarían a predicciones incorrectas en estudios *in silico*. La asignación de cada gen-proteína a una reacción fue validada por la realización de BLASTP de cada proteína frente a la base de datos de Swiss-Prot (The UniProt Consortium, 2007) y la de *Pseudomonas* (<http://www.pseudomonas.com/blast/set>). Las proteínas de TFA se consideraron ortólogas con el mejor resultado del alineamiento si el e-value era menor de  $1e-5$ , la identidad mayor del 30 % y la cobertura de ambas secuencias mayor del 70 %. Además, se accedió a bases de datos bioquímicas y genómicas, tales como KEGG (Kanehisa y Goto, 2000), BIGG (Schellenberger y col., 2010), BRENDA (Schomburg y col., 2002) y MetaCyc (Karp y col., 2002), y a la literatura científica para buscar información sobre las reacciones y las enzimas que las llevan a cabo. De este modo, se validaron o descartaron las reacciones existentes en el modelo inicial y añadieron nuevas GPR que debían estar presentes en el metabolismo de TFA y no habían sido incorporadas al borrador inicial.

### Asignación del índice de confianza a cada reacción

Durante la curación manual, a cada reacción del modelo se le asignó un índice de confianza (valores entre 1 y 4) el cual indica el grado de conocimiento biológico que avala su inclusión en el modelo según lo indicado en la Tabla 6. Este índice resultó de gran utilidad a la hora de eliminar o mantener una reacción en el proceso de validación del modelo. Además, cuando fue necesario, se incluyeron notas y referencias bibliográficas a las GPRs, que igualmente permitieron evaluar la inclusión de la reacción en el modelo.

**Tabla 6. Criterios para indicar el nivel de fiabilidad de la inclusión de una reacción en el modelo metabólico**

<b>Evidencia</b>	<b>Nivel de confianza</b>	<b>Ejemplo</b>
Datos bioquímicos	4	Evidencias directas para las funciones de productos de genes y reacciones bioquímicas: purificación de proteínas, ensayos bioquímicos.
Datos genéticos	3	Evidencias directas e indirectas para la función de un gen: caracterización de mutantes, sobreexpresión.
Datos fisiológicos	2	Evidencias indirectas para reacciones bioquímicas basadas en datos fisiológicos: productos de secreción o componentes definidos en el medio que sirven como evidencia para el transporte y reacciones metabólicas. Evidencias para la función del gen basadas en la anotación.
Datos del modelo	1	No se encuentran evidencias pero la reacción es necesaria para el funcionamiento del modelo.

### Adición de rutas metabólicas específicas de TFA

A pesar de que el modelo inicial incorpora con alta fidelidad las reacciones del metabolismo central y otras ampliamente conservadas en bacterias, no ocurre lo mismo con las reacciones específicas del metabolismo que se está reconstruyendo. En el caso de TFA, se utilizó el conocimiento fisiológico y la anotación de la bacteria para reconstruir rutas metabólicas no englobadas en el modelo inicial.

### Inclusión en el modelo de reacciones sin gen-proteína asociados

El modelo metabólico requiere de determinadas reacciones que, por definición, no tienen asignada una proteína. Se trata de las reacciones de intercambio de metabolitos con el exterior, que suelen ser nombradas como “EX\_N(e)”, donde “N” es el nombre abreviado del metabolito y “e” hace referencia al exterior celular. Además de estas, en el modelo de TFA se han incluido otras reacciones sin gen-proteína asociados como las de demanda de



metabolitos y las de mantenimiento, siendo un ejemplo la del requerimiento de ATP no asociado al crecimiento (ATPM, del inglés *ATP maintenance requirement*). Del mismo modo, las reacciones que ocurren de forma espontánea o las reacciones de transporte por difusión tampoco llevan asignado un gen.

Por otra parte, cuando en la anotación de TFA no se identificaron las proteínas requeridas para llevar a cabo reacciones bioquímicas bien conocidas, se incluyeron dichas reacciones en el modelo como reacciones huérfanas, es decir, sin asociación gen-proteína. Entre ellas se encuentran reacciones de transporte de metabolitos para los que existen evidencias de su transporte al interior celular a pesar de no haber identificado un transportador.

### Construcción de la reacción de biomasa

Para evaluar si el modelo era capaz de generar biomasa *in silico*, es decir, de emular el crecimiento del microorganismo, se definió una reacción de biomasa que se usaría como función objetivo (BOF, del inglés *Biomass Objective Function*). Esta reacción contiene los metabolitos que componen la biomasa celular y en cada uno de ellos se indica el coeficiente con el cual contribuyen al total de la biomasa. Debido a que actualmente la información sobre la composición macromolecular de *Sphingopyxis granuli* TFA es insuficiente para construir una BOF específica, se asumió una BOF típica de bacterias Gram negativas basada en la de *P. putida* (Nogales y col., 2008), con ciertas modificaciones basadas en la información disponible de la fisiología de TFA (Tabla 7).

**Tabla 7. Composición de la reacción de biomasa de TFA**

Categorías	Componentes
Lípidos y componentes de la pared celular	Mureína, esfinganina, fosfatidiletanolamina, fosfatidilglicerol, cardiolipina.
Aminoácidos	L-aminoácidos (alanina, glutamato, glicina, lisina, aspartato, arginina, glutamina, serina, metionina, triptófano, fenilalanina, tirosina, cisteína, leucina, histidina, prolina, asparagina, valina, treonina, isoleucina.
Cofactores y vitaminas	Coenzima A (CoA) y sus derivados acetil-CoA, succinil-CoA, y malonil-CoA; ubiquinol, riboflavina, N5 y N10-formil tetrahidrofolato, 5,10-metilentetrahidrofolato, biotinil-5'-AMP, hemo, piridoxal-5-fosfato, sirohemo, tiamina difosfato
Nucleótidos y moléculas energéticas	dATP, dTTP, dGTP, dCTP, ATP, UTP, GTP, CTP, NAD, NADH, NADPH, NADP, FAD, AMP, dinucleótido de molibdopterina y guanina
Otros	Iones inorgánicos, glutatión reducido, lipoato, 5-hidroxiectoina, corismato, putrescina, espermidina

### **3.2.3. Conversión del modelo a formato computacional**

En la tercera etapa se realizó la conversión del modelo, hasta el momento en formato Excel, a formato computacional. Para ello, se utilizó el paquete COBRA (<https://opencobra.github.io/cobratoolbox/>) a través del entorno de computación Matlab y se ejecutó el siguiente comando que llama a la función xls2model:

```
model = xls2model('S_granuli_TFA.xls')
```

De esta forma, el modelo quedó disponible en Matlab para los siguientes análisis, los cuales se llevaron a cabo gracias a los programas incluidos en el paquete COBRA.

### **3.2.4. Evaluación inicial y relleno de huecos (*gap filling*) en las rutas metabólicas del modelo**

#### Selección de la función objetivo y FBA

Para resolver el modelo, es decir, encontrar una solución a los problemas de programación lineal, se usó la función FBA (*Flux Balance Analysis*). Para ello, se fijó una función objetivo que, en el caso de evaluar el crecimiento del modelo, fue la reacción de biomasa (BOF).

Posteriormente, para optimizar dicha función objetivo, el sistema de ecuaciones lineales se resolvió con la aproximación FBA. En este análisis se asume la condición de estado estacionario, en el que la cantidad de compuestos producidos tiene que ser igual al total de metabolitos consumidos. De esta forma, se obtuvieron los flujos metabólicos que maximizaban la tasa de crecimiento, en unidades de mmol/g (peso seco)/h.

#### Validación del modelo y *gap filling*

Para llevar a cabo la validación del modelo, se comprobó que todas las reacciones estuviesen balanceadas y que todos los metabolitos presentes en las reacciones del modelo estuviesen definidos junto con su carga y asignados al compartimento adecuado (citoplasma, periplasma y/o espacio extracelular).

Además, se evaluó si todos los componentes de la reacción de biomasa podían ser sintetizados a través de la red de reacciones del modelo, partiendo de un medio mínimo definido y una de las fuentes de carbono (tetralina, ácido sebácico o  $\beta$ -hidroxibutirato) en las que el crecimiento *in vivo* de TFA ha sido ampliamente caracterizado. En el caso de que

algunos metabolitos de la biomasa no se sintetizaran, se evaluaron una a una las reacciones que conducían a su producción con el fin de rellenar los huecos (reacciones no incluidas) en las distintas rutas metabólicas.

Del mismo modo, se analizaron las rutas metabólicas de los distintos subsistemas, como la biosíntesis y metabolismo de aminoácidos, de lípidos, de purinas y pirimidinas, la síntesis de cofactores, etc. Análogamente a lo mencionado en el párrafo anterior, cuando se encontraron huecos en las rutas, se introdujeron las reacciones que faltaban en el modelo. A estas reacciones que rellenaban los huecos en las rutas metabólicas se les asignaron genes cuando fue posible, se dejaron huérfanas o se indicó, en su caso, que se trataban de reacciones espontáneas según la información disponible en las bases de datos y en la bibliografía.

### **3.2.5. Denominación del modelo**

El modelo se nombró según las reglas de nomenclatura establecidas en Reed y col. (2003). De esta forma, el modelo metabólico de *S. granuli* TFA se denominó *i*IG738, donde *i* hace referencia a *in silico*, IG son las siglas de la persona que construyó el modelo (Inmaculada García) y 738 indica el número de genes presentes en la reconstrucción.

### **3.2.6. Evaluación de la capacidad predictiva del modelo**

#### Validación cualitativa

Para determinar la capacidad predictiva del modelo *i*IG738, se realizó una predicción *in silico* del crecimiento de TFA definiendo un medio mínimo con distintas fuentes de carbono y energía. Dichas fuente fueron elegidas en base a los datos de la anotación del genoma y del propio modelo, cuando había indicios de que fuesen sustratos con los que TFA podría crecer.

#### Validación cuantitativa

El modelo requiere de la tasa de consumo real de la fuente de carbono para predecir la velocidad de crecimiento. La tasa de entrada del resto de metabolitos del medio mínimo se fija a -30 mmol/g (peso seco)/h.

Para la medición del consumo de la fuente de carbono, se iniciaron cultivos de 20 mL en MM a una D.O.<sub>600</sub> de 0,1 a partir de inóculos en fase exponencial crecidos en MM con la misma fuente de carbono que el nuevo medio a inocular. Se seleccionaron dos fuentes de carbono,  $\beta$ -hidroxibutirato 40 mM y ácido sebácico 16 mM, en las que TFA presentaba una tasa de

## ***Materiales y Métodos***

crecimiento mayor, y otras dos energéticamente menos favorables, como son L-prolina 32 mM y L-lactato 53 mM, ajustándose la concentración para que el número de carbonos fuese el mismo entre cada fuente de carbono. A diferentes D.O.<sub>600</sub>, se recogió 1 mL de cultivo al inicio de la curva, se centrifugó 5 minutos a velocidad máxima y se almacenó el sobrenadante a -20 °C .

La medida de la cantidad de fuente de carbono remanente en el sobrenadante de las muestras se realizó mediante GC-MS (cromatografía de gases y espectrometría de masas). Para ello, 100 µL de sobrenadante se evaporaron en seco con nitrógeno a 50 °C en un tubo de fondo cónico. A continuación, el precipitado se resuspendió en 100 µL de una solución de 0,5 mg/mL de ácido cítrico en piridina. Seguidamente, se añadieron 200 µL de BSTFA+TMCS (99:1), la mezcla se agitó vigorosamente durante 5 segundos y se dejó reaccionando durante 45 minutos a 100 °C. Finalmente, 1 µL de muestra se inyectó en una columna DB-17 del sistema GC-MS y se obtuvo el área de las señales correspondiente a la fuente de carbono y al citrato, control interno con el cual se normalizó la señal del sustrato.

Previamente a la medida del sustrato en el sobrenadante, se realizó una recta de calibrado empleando muestras de concentración conocida del compuesto en MM. De esta forma, se extrajo la ecuación lineal que relacionaba el área normalizada del metabolito con su concentración. A partir de dicha recta de calibrado, se calculó la concentración en mM de la fuente de carbono en el sobrenadante de los cultivos.

Los flujos en el modelo metabólico se deben expresar en mmol/g (peso seco)/h. Para conseguir el valor de consumo en estas unidades, se calculó el peso seco de cultivos de TFA en cada una de las fuentes de carbono analizadas. En cada caso, se tomaron 100 mL de cultivo a distintos tiempos de la curva de crecimiento, se centrifugaron a 5000 rpm durante 15 minutos, se eliminó el sobrenadante y el sedimento se resuspendió en 2 mL de tampón fosfato 1X y se transfirió a un tubo Eppendorf de 2 mL. A continuación, los tubos se centrifugaron 5 minutos a velocidad máxima en una microcentrífuga y se eliminó el sobrenadante. Finalmente, el precipitado se secó a 80°C durante el tiempo necesario para que se despegase del fondo del Eppendorf y se tornara color marrón, momento en el que se pesó en la balanza de precisión.

Tampón fosfato 50X: Na<sub>2</sub>HPO<sub>4</sub>·12H<sub>2</sub>O 17 mM, KH<sub>2</sub>PO<sub>4</sub> 7,35 mM. Autoclavar

#### 4. Herramientas bioinformáticas para los análisis de genómica comparada

Los genomas y proteomas que se han utilizado a lo largo de esta tesis doctoral para realizar los análisis de genómica comparada se listan en el Anexo II.

##### 4.1. Construcción de árboles filogenéticos

Para la adscripción filogenética de TFA se utilizó el servicio de identificación incorporado en EzBioCloud (<https://www.ezbiocloud.net/>) (Yoon y col., 2017), que compara la secuencia del gen codificante para el ARNr 16S con una extensa base de datos de ADNr 16S de bacterias cultivadas y no cultivadas. Posteriormente se realizaron distintos árboles filogenéticos comparando las secuencias, de genes concretos o de proteomas completos, de representantes de los géneros de la familia *Sphingomonadaceae*.

##### 4.1.1. Concatenación de genes conservados

Se concatenaron las secuencias de genes conservados ampliamente utilizados en filogenia, como son el gen 16S ARNr, *atpD* (subunidad beta de la ATP sintasa) y *rpoB* (subunidad beta de la polimerasa de ARN). Posteriormente, los genes concatenados de las distintas bacterias se alinearon mediante Clustal Omega y dicho alineamiento se utilizó para construir la filogenia aplicando el método Neighbor-Joining a través de la herramienta SeaView (Gouy y col., 2010). Se aplicó un bootstrap de 1000 repeticiones.

##### 4.1.2. Cálculo de la media en identidad de aminoácidos (AAI, del inglés *Average Amino Acid Identity*)

Para calcular la AAI se alinearon una a una todas las secuencias proteicas codificadas en un genoma frente al proteoma completo de otra bacteria, repitiendo el proceso con el resto de los proteomas hasta obtener todas las comparaciones posibles. Para tal fin se utilizó una versión modificada del módulo 2 de Sma3s (basado en BLASTP recíproco, ver Anexo III) que dio como resultado una lista de porcentajes de identidad para cada comparación. En el análisis, sólo se consideraron secuencias ortólogas aquellas cuyos alineamientos cumplieran con una identidad >30 %, cobertura de ambas secuencias de >70 % y e-value  $\leq 1e-05$ , como se recomienda en Konstantinidis y Tiedje (2005).

Los valores de identidad entre proteomas fueron utilizados para construir una matriz de distancia y generar un dendograma mediante el programa fneighbor del paquete embassy-phyliip, aplicando el método Neighbor-Joining (Saitou y Nei, 1987) y visualizando el dendograma con SeaView.

### **4.1.3. Cálculo de la media en identidad de nucleótidos (ANI, del inglés *Average Nucleotide Identity*)**

Para obtener la media en identidad de nucleótidos entre los genomas de TFA y especies cercanas se utilizó la herramienta ANIm, que se basa en una comparación por MUMmer de todos los genomas y forma parte de las herramientas disponibles en el servidor JSpeciesWS (Richter y col., 2015).

### **4.2. Comparación de secuencias genómicas y contexto génico**

La comparación de la secuencia del genoma de TFA frente al de otras bacterias se realizó mediante BLASTN con la herramienta BRIG (Alikhan y col., 2011). Los ficheros de entrada en dicho programa consistieron en los genomas de las bacterias comparadas (Anexo II) en formato FASTA. Con esta herramienta también se analizó el contenido en GC a lo largo del genoma manteniendo los parámetros por defecto. Además, se utilizó el programa DNAplotter v1.11 (Carver y col., 2009) para representar la desviación GC a lo largo del genoma de TFA, calculada como  $(G-C)/(G+C)$  y utilizando una ventana de 10000 nucleótidos con un desplazamiento de 200 nucleótidos hasta recorrer la secuencia completa del genoma. En esta herramienta, el fichero de entrada también fue la secuencia del genoma de TFA en formato FASTA.

Para comparar el contexto génico de un gen de interés entre distintos genomas bacterianos se utilizó el servidor web SyntTax (Oberto, 2013). Como entrada se introdujo la secuencia proteica codificada por el gen de interés y se seleccionó el conjunto de bacterias en las que se quería analizar el contexto génico. Los parámetros se mantuvieron por defecto.

### **4.3. Construcción de una matriz de proteomas**

Para la construcción de la matriz piramidal en la que se comparan entre sí todos los proteomas se utilizó la herramienta matrix del paquete CMG-biotools (Vesth y col., 2013). Dicho paquete se distribuye como una máquina virtual de Linux la cual contiene numerosos scripts útiles para analizar genomas bacterianos y realizar genómica comparada. Los parámetros que el paquete sigue para considerar dos proteínas ortólogas se modificaron en el fichero *tigr\_cut* para establecer un mínimo de identidad del 30 % y un mínimo de cobertura del 70 % de la secuencia más larga del alineamiento. Una vez establecidos los parámetros, se generó una carpeta conteniendo los ficheros .fsa (proteínas en formato FASTA) de todos los proteomas a analizar. A continuación, se ejecutó el programa *matrix\_createConfig* para construir un fichero en formato XML con los datos necesarios para construir la matriz, el cual

se usó como fichero de entrada para el script *matrix*. Finalmente se obtuvo una figura en formato .ps que correspondía a una representación en pirámide de la comparación entre proteomas.

#### 4.4. Obtención del pangenoma y el genoma núcleo del género *Sphingopyxis*

Para la construcción del pangenoma y el genoma núcleo del género *Sphingopyxis* se empleó el script *pancoreplot* del paquete CMG-biotools, teniendo en cuenta los mismos parámetros que en el apartado anterior para considerar ortólogas dos proteínas. Del mismo modo que en el apartado anterior, se colocaron todos los ficheros .fsa en una misma carpeta y se ejecutó el programa *pancoreplot\_createConfig* para generar un fichero con dos columnas: la primera conteniendo el nombre de la bacteria tal y como aparecerá en la gráfica y la segunda el nombre del fichero que contiene el proteoma de dicha bacteria. Este archivo se usó como fichero de entrada para el script *pancoreplot* el cual generó un gráfico representativo del pangenoma y genoma núcleo y una carpeta que guarda los datos crudos de salida del programa.

Finalmente, para extraer las proteínas del genoma núcleo, comunes a todas las *Sphingopyxis* analizadas, se ejecutó la herramienta *pancoreplot\_subsets* con los argumentos - o 1 y - i 1, tomando como input los datos crudos generados previamente por *pancoreplot*. Los genes específicos de TFA se extrajeron con la misma herramienta pero estableciendo los argumentos - o 1 - i 1 -c 2:9.

### 5. Transferencia de ADN a estirpes bacterianas

#### 5.1. Transformación

##### 5.1.1. Transformación de *E. coli* mediante choque térmico

##### Preparación de células competentes

*Método 1.* Para transformar células de *E. coli* DH5 $\alpha$  con plásmidos o mezclas de ligaciones, se prepararon células competentes mediante el protocolo establecido por Inoue y colaboradores (1990), con el que se consigue una alta frecuencia de transformación (de hasta 10<sup>9</sup> transformantes por cada  $\mu$ g de ADN plasmídico). Para ello, un cultivo saturado de DH5 $\alpha$  se diluyó 100 veces en 200 mL de medio SOB y se incubó a 22 °C en agitación hasta una D.O.<sub>600</sub> de 0,5. A continuación, el cultivo se colocó rápidamente en hielo donde se mantuvo durante 10 minutos. Posteriormente se centrifugó el cultivo a 2500 g durante 10 minutos a 4 °C, se eliminó el sobrenadante y el sedimento de células se resuspendió en 20 mL de TB frío, a los que se añadió 60 mL de nuevo TB frío. Tras una incubación en hielo de 10 minutos se

## ***Materiales y Métodos***

volvieron a centrifugar en las mismas condiciones, resuspendiéndose en 20 mL de TB frío. Por último, se añadieron 1,5 mL de DMSO, se incubaron 10 minutos en hielo y se hicieron alícuotas de 0,3 y 0,5 mL que se congelaron rápidamente en nitrógeno líquido y se almacenaron a -80 °C.

Medio SOB: triptona 2 % (p/v), extracto de levadura 0,5 % (p/v), NaCl 8,5 mM, KCl 1,25 mM. Ajustar el pH a 7,0 con NaOH y autoclavar. Añadir posteriormente MgCl<sub>2</sub> 10 mM y glucosa 0,36 mM.

TB: PIPES (ácido libre) 10 mM, CaCl<sub>2</sub>·2H<sub>2</sub>O 15 mM, KCl 250 mM. Ajustar el pH a 6,7 con KOH. Añadir MnCl<sub>2</sub>·4H<sub>2</sub>O 55 mM. Esterilizar por filtración.

**Método 2.** Alternativamente, para la transformación con plásmidos de estirpes de *E. coli*, las células se hicieron competentes mediante un tratamiento químico con TSS (Chung y col., 1989), el cual requiere que las células se transformen inmediatamente puesto que no se pueden conservar. Inicialmente, un cultivo saturado de la estirpe receptora se diluyó 100 veces en medio LB y se incubó a 37 °C hasta alcanzar la fase exponencial temprana. En este punto, el cultivo se repartió en alícuotas de 1 mL en tubos Eppendorf que se enfriaron en hielo durante 5 minutos y se centrifugaron 30 segundos a velocidad máxima en una microcentrífuga. Posteriormente, se retiró el sobrenadante, se añadieron 75 µL de LB frío, se incubaron 5 minutos en hielo y, por último, se añadieron 75 µL de TSS 2X frío. Las muestras se agitaron suavemente y se volvieron a incubar 5 minutos en hielo, obteniendo así células preparadas para ser transformadas.

TSS 2X: triptona 0,8 % (p/v), extracto de levadura 0,5 % (p/v), NaCl 85,5 mM, PEG-8000 25 mM, DMSO 10 % (v/v), MgSO<sub>4</sub> 100 mM, pH 6,5.

### **Transformación**

Para la transformación se añadieron entre 1 y 5 µL de una preparación de plásmidos o, cuando correspondía, 10 µL de una reacción de ligación a 100 µL de células competentes. Dicha mezcla se incubó 30 minutos en hielo y, a continuación, se sometió a un choque térmico a 42 °C durante 45 segundos. Posteriormente, las células se pasaron rápidamente a hielo y se incubaron 5 minutos. Tras ese tiempo, se les añadió 1 mL de SOC y se incubaron a 37 °C durante 1h y a 180 rpm de agitación. Finalmente, se sembraron en placas de medio selectivo y se incubaron toda la noche a 37 °C.

#### **5.1.2. Electrotransformación de *S. granuli* TFA**

##### **Preparación de células electrocompetentes**

Para la preparación de células electrocompetentes de *S. granuli* se inocularon inicialmente 10 mL de MML con una colonia de la estirpe receptora y se incubaron a 30 °C durante aproximadamente 24 horas, hasta alcanzar una D.O.<sub>600</sub> entre 1,5 y 2 (saturación).



Este inóculo se diluyó entonces a D.O.<sub>600</sub> de 0,1 en 100 mL de MML fresco que se volvió a incubar a 30 °C hasta alcanzar una D.O.<sub>600</sub> de aproximadamente 0,4. Las siguientes etapas del proceso se realizaron manteniendo siempre las células en frío. El cultivo se enfrió en hielo durante 10 minutos y se recogieron las células mediante centrifugación a 5000 rpm durante 10 minutos a 4 °C. A continuación, las células se lavaron sucesivamente en 100 mL y 50 mL de agua bidestilada y 2 mL de glicerol al 10 % en agua bidestilada (v/v). Por último, las células se resuspendieron en 250 µL de glicerol al 10 % y se guardaron alícuotas de 40 µL a -80 °C.

#### Electroporación de células electrocompetentes de TFA

A cada alícuota de electrocompetentes se le añadieron unos 50 ng de ADN plasmídico libre de sales en un volumen máximo de 4 µL, manteniéndolas en hielo. Las mezclas se introdujeron en cubetas de electroporación con una distancia entre electrodos de 2 mm y se les aplicó una diferencia de potencial de 2,5 kV en un electroporador Micropulser TM (Bio-Rad). Inmediatamente después se agregó a cada cubeta 1 mL de medio frío MML con sorbitol 0,5 M y se transfirió todo el volumen a tubos Eppendorf estériles, que se incubaron 60 minutos a 30 °C. Finalmente se sembraron en medio sólido selectivo y se incubaron 5 ó 6 días a la misma temperatura.

### **5.2. Conjugación triparental**

Para transferir plásmidos movilizables desde *E. coli* (estirpe donadora) a *S. granuli* TFA (estirpe receptora), se utilizó como cepa auxiliar *E. coli* DH5α transformada con el plásmido pRK2013, el cual porta los genes de transferencia (Figurski y Helinski, 1979). Tanto la estirpe de *E. coli* donadora como la auxiliar se cultivaron en LB a 37 °C, mientras que la estirpe receptora de TFA se incubó a 30 °C en MML, todas ellas hasta obtener cultivos saturados. Posteriormente, los pre-inóculos se diluyeron a una D.O.<sub>600</sub> de 0,2 en medio rico sin antibióticos y se cultivaron hasta alcanzar una D.O.<sub>600</sub> de 0,4. Como TFA crece más lento que *E. coli*, la dilución de la estirpe receptora se realizó 1 hora y 30 minutos antes que la donadora y la auxiliar. A continuación, se mezclaron los cultivos en proporción 1:1:5 (donador:auxiliar:receptor) en un volumen final de 1 mL. La mezcla se centrifugó 10 minutos a 13.000 rpm y el sedimento se resuspendió en 100 µL de tampón fosfato 1X, los cuales se depositaron sobre el centro de una caja de MML que se incubó a 30 °C durante toda la noche. Tras la incubación, se recogió la biomasa crecida con un asa de siembra y se resuspendió en 1 mL de tampón fosfato 1X. Finalmente, las células se sembraron en medio selectivo a distintas diluciones, normalmente de 10<sup>1</sup> a 10<sup>5</sup>, y se incubaron a 30 °C durante 5-7 días hasta la detección de colonias visibles.

## **6. Manipulación de ácidos nucleicos**

### **6.1. Manipulación de ADN**

#### **6.1.1. Aislamiento de ADN plasmídico de *Escherichia coli***

Para aislar ADN plasmídico de *E. coli* DH5 $\alpha$  se siguió el protocolo simplificado de lisis alcalina (Stephen y col., 1990). La biomasa de 3 mL de cultivo saturado se recogió por centrifugación a 16.100 g durante 1 minutos y se resuspendió en 100  $\mu$ L de GTE frío, con la posterior incubación 5 minutos en hielo. A la solución anterior se le añadieron 200  $\mu$ L de solución II recién preparada, se homogeneizó la muestra por inversión y la mezcla formada se volvió a incubar 5 minutos en hielo. A continuación, se añadieron 150  $\mu$ L de solución III y se procedió a homogeneizar e incubar la muestra como en el paso anterior. Los restos celulares se sedimentaron por centrifugación 5 minutos a 16.100 g y el sobrenadante se transfirió a un nuevo tubo al que se añadió 1 volumen (450  $\mu$ L) de etanol al 96 % (v/v) a -20 °C. Inmediatamente después, la muestra se centrifugó 1 minuto a 16.100 g y el precipitado se lavó con 1 mL de etanol al 70 % (v/v) a -20 °C, se secó mediante vacío y se resuspendió en 25 ó 50  $\mu$ L de TER, dependiendo del número de copias del plásmido por célula.

GTE: glucosa 50 mM, EDTA 10 M, Tris-HCl 25 mM pH 8.

Solución II: NaOH 0,2N, SDS 1 % (p/v). Preparar fresca en el momento de usar.

Solución III: acetato potásico 3 M a pH ajustado a pH 4,8 con ácido fórmico.

TER: Tris-HCl 10 mM pH 8, EDTA 1 mM, ribonucleasa (Roche) 0,02 % (p/v).

Cuando se requirió ADN plasmídico de alta calidad y mayor pureza, se usó el kit Nucleospin® (Marchery-Nagel) siguiendo el protocolo establecido por el fabricante.

#### **6.1.2. Aislamiento de ADN total de *S. granuli***

Las extracciones de ADN total de *S. granuli* se llevaron a cabo a partir de las células de 3 mL de cultivo saturado crecido en MML a 30 °C utilizando el kit Wizard® Genomic DNA purification (Promega) según las instrucciones del fabricante.

#### **6.1.3. Electroforesis de ADN en geles de agarosa**

Para evaluar la calidad y el tamaño del ADN procedente de extracciones, fragmentos de reacciones de PCR, fragmentos de reacciones de restricción u otra procedencia se empleó la electroforesis convencional en gel de agarosa. Normalmente se utilizó agarosa de baja electroendosmosis disuelta en tampón TAE 1X a concentraciones entre 0,6 y 2,5 % (p/v), dependiendo del tamaño de los fragmentos a separar. Antes de cargar las muestras en los

pocillos del gel, se les añadió tampón de carga 6X y como patrón de peso molecular se usó el marcador 1 kb ladder (Invitrogen). Los geles se corrieron en tampón TAE 1X aplicando una diferencia de potencial que varió entre 45 y 135 V.

Para visualizar el ADN tras la electroforesis, los geles se tiñeron en una solución de bromuro de etidio a 2,5 mg/L. Posteriormente se irradiaron con un transiluminador de radiación ultravioleta (Vilber Lourmat) y las imágenes se obtuvieron con una cámara acoplada a una impresora térmica.

Tampón de carga 6X: azul de bromofenol 0,05 % (p/v), xileno de cianol 0,05 % (p/v), glicerol 30 % (v/v), TE 50X 2 % (v/v).

TAE 1X: Tris-ácido acético 40 mM, EDTA 10 mM, pH 7,7.

#### **6.1.4. Purificación de fragmentos de ADN**

Los fragmentos de ADN procedentes de reacciones enzimáticas o separados en geles de agarosa se purificaron con el kit GFX (GE healthcare) según las instrucciones del fabricante. En el caso de la purificación de fragmentos de ADN desde gel de agarosa, se cortó con un bisturí la región del gel que contenía el fragmento de interés y se procedió a su extracción con el protocolo especificado en el mismo kit.

#### **6.1.5. Clonación de fragmentos de ADN**

Tanto los insertos procedentes de plásmidos o reacciones de PCR, como los vectores para la clonación se digirieron con endonucleasas de restricción adquiridas de las compañías Roche o New England Biolabs. En cada caso se siguieron las recomendaciones del proveedor y se empleó el tampón de restricción suministrado por la casa comercial junto a la enzima de restricción.

Las reacciones de ligación se realizaron en un volumen final de 10 µL en tampón quinasa con ATP a una concentración de 1 mM, cuando los fragmentos a ligar presentaban al menos un extremo cohesivo, o de 0,1 mM, cuando los extremos a ligar eran romos. A cada reacción se añadió 1 unidad de ADN ligasa de T4 (Roche) y se incubó un mínimo de 12 horas a 16 °C.

Tampón quinasa 10X: Tris-HCl 0,5 M pH 7,6, Cl<sub>2</sub>Mg 0,1 M, DTT 50 mM, BSA 0,05 % (p/v)

#### **6.1.6. Detección de secuencias de ADN por hibridación en membrana (*Southern Blot*)**

Para la realización de *Southern Blot* se diseñaron sondas de ADN homólogas a la secuencia del fragmento que se quería detectar. Las sondas empleadas consistieron en fragmentos de restricción o de PCR aislados de geles de agarosa. El marcaje de las mismas se realizó mediante el método de cebado aleatorio o *random priming* con digoxigenina-dUTP (deoxiuridín nucleósido trifosfato, Roche) según las instrucciones del proveedor.

Un total de 2 µg del ADN de cada muestra se digirió con las endonucleasas de restricción convenientes. Para separar los fragmentos de la digestión se llevó a cabo una electroforesis convencional en gel de agarosa al 0,6 % (p/v), a 45 V y 4 °C. El posterior tratamiento de los geles se realizó según las indicaciones de Sambrook y colaboradores (Sambrook y col., 1989). En primer lugar, los geles se incubaron en las soluciones de despurinización (20 min), desnaturalización (30 min) y neutralización (2 x 15 min). A continuación, el ADN se transfirió desde el gel a una membrana de nailon Hybond N (GE Healthcare) mediante vacío durante 2 horas utilizando SSC 10X como tampón de transferencia. Una vez transferido el ADN, la membrana se secó en la campana de flujo laminar y el ADN se fijó a la misma a través de la aplicación de luz ultravioleta a 254 nm de longitud de onda durante 30 segundos.

La membrana se prehibridó un mínimo de 1 h a 42 °C en un horno de hibridación Techne® HB-3D con solución de prehibridación precalentada 10 minutos a 65 °C. La hibridación posterior se llevó a cabo durante unas 14-16 horas a 42 °C en solución de prehibridación a la que se añadió la sonda previamente hervida. A continuación, se hicieron dos lavados de 5 minutos cada uno a temperatura ambiente con SSC 2X, SDS 0,1 % (p/v) y posteriormente dos lavados más de 15 minutos cada uno con SSC 0,1X, SDS 0,1 % (p/v) a 68 °C.

La detección de la sonda marcada se llevó a cabo mediante un anticuerpo anti-digoxigenina-dUTP conjugado con fosfatasa alcalina (Roche). Todos los pasos siguientes se realizaron a temperatura ambiente y en agitación. La membrana se equilibró 1 minuto en Tampón 1. Seguidamente se bloqueó durante al menos 1 hora en solución de bloqueo, tiempo tras el cual se añadió el anticuerpo, en una proporción de 1 µL por cada 10 mL de solución de bloqueo, y se incubó durante 30 minutos. Después se eliminó la solución con el anticuerpo y se lavó la membrana dos veces con Tween 20 al 0,3 % (v/v) en Tampón 1. Una vez lavada la membrana, se equilibró 1 minuto en Tampón 3. A continuación, se incubó en una solución 1:300 de CSPD® (Roche), sustrato luminiscente de la fosfatasa alcalina. Finalmente, las señales luminiscentes se detectaron a través del sistema de imagen ChemiDoc™ (BioRad).

Solución de depurinización: HCl 0,25 M.

Solución de desnaturalización: NaCl 1,5 M, NaOH 0,5 M.

Solución de neutralización: Tris-HCl 1 M, NaCl 1,5 M, pH 7,4.

SSC 20X: NaCl 175 g/L, Citrato sódico  $\cdot 2H_2O$  88 g/L, ajustar a pH 7 con HCl 1 M.

Solución de prehibridación: SSC 5X, N-laurilsarcosina 0,1 % (p/v), blocking reagent 1 % (p/v) (Roche), formamida 50 % (v/v), ADN de esperma de salmón 0,05 mg/L. El ADN de esperma de salmón debe añadirse en el momento de usar la solución, previamente hervido durante 10 minutos.

Tampón 1: Ácido Maleico 100 mM, NaCl 150 mM, pH 7,5.

Solución de bloqueo: Blocking reagent (Roche) al 10 % (p/v) en Tampón 1.

#### **6.1.7. Detección de secuencias de ADN por hibridación en colonias**

Las colonias procedentes de una genoteca ordenada de TFA (construida previamente en el laboratorio), se crecieron toda la noche a 37 °C en placas de LB con ampicilina. Posteriormente, se recortaron membranas de nailon Hybond N (GE Healthcare) del diámetro de las placas y se depositaron sobre las colonias unos 45 segundos. Una vez transferidas todas las colonias a las membranas, estas se incubaron de forma secuencial en las siguientes soluciones durante 3 minutos: (1) SDS 10 %, (2) NaCl 1,5M + NaOH 0,5M, (3) dos veces en Tris 1M + NaCl 1,5M y (4) SSC 2X (solo unos segundos para lavarlas). Entre cada paso, se secan ligeramente en papel Whatman. Finalmente, las membranas se dejaron secar durante 1h a temperatura ambiente y se fijó la transferencia incubándolas 30 segundos por cada lado con luz ultravioleta. Para los siguientes pasos de prehibridación, hibridación y detección se procedió como en la sección anterior.

#### **6.1.8. Reacción en cadena de la polimerasa (PCR)**

Las reacciones en cadena de la polimerasa para amplificar fragmentos de ADN de interés se realizaron empleando distintas polimerasas de ADN según la calidad requerida en el producto de PCR. Por tanto, para obtener fragmentos de alta fidelidad para clonaciones o para el ensamblaje del genoma de TFA se usaron las enzimas iPfu Polymerase (Thermo Scientific) o Expand<sup>TM</sup> High Fidelity (Roche). Para reacciones de comprobación en las que no se requirió secuenciar los fragmentos obtenidos y no se necesitó una alta fidelidad se usaron las polimerasas Taq DNA Polymerase (Promega) o DreamTaq (Thermo). En ciertas ocasiones se usó el kit Ilustra<sup>TM</sup> PureTaq<sup>TM</sup> Ready-to-Go<sup>TM</sup> PCR beads (GE Healthcare).

De forma general se usaron 25 ng de ADN molde en un volumen final de reacción de 20, 25 ó 50  $\mu$ L, las cuales se prepararon en tubos de 0,2 mL. A veces se usó como molde ADN procedente de células hervidas: en el caso de *E. coli*, la biomasa procedente de una colonia se depositó en la mezcla de PCR previamente preparada mientras que para TFA, la colonia

se resuspendió en 30 ó 50  $\mu\text{L}$  y se hirvió 10 minutos a 100 °C. Tras este tiempo, la mezcla se centrifugó y se usaron 5  $\mu\text{L}$  del sobrenadante como molde para la amplificación. Los oligonucleótidos se añadieron a una concentración final de 1  $\mu\text{M}$ , los dNTPs a una concentración final de 50  $\mu\text{M}$  y de enzima se utilizaron de 2 a 4 unidades por reacción, todo ello en el tampón proporcionado por la casa comercial (con el  $\text{Mg}^{2+}$  necesario ya incorporado). Finalmente, en muchas de las reacciones fue necesario añadir DMSO a una concentración final de 5 % (v/v), lo cual facilitó la amplificación de regiones con alto contenido en G+C (Jensen y col., 2010).

Las reacciones se realizaron en un termociclador Biometra Tpersonal, empleando el programa adecuado según la temperatura óptima de hibridación de los oligonucleótidos, el tamaño del fragmento a amplificar y las especificaciones de la casa comercial de la enzima polimerasa. Cuando fue necesario, se llevaron a cabo PCRs en gradiente de temperatura para optimizar la hibridación de los oligonucleótidos y la amplificación del fragmento deseado.

### **6.1.9. Hibridación y extensión de oligonucleótidos solapantes**

En ciertos casos, para obtener secuencias de ADN con regiones mutadas con respecto a la secuencia silvestre, se sintetizaron dos oligonucleótidos parcialmente complementarios, que juntos cubrían toda la región de interés. La región solapante se hacía coincidir con la zona en la que se querían introducir las mutaciones, las cuales se incorporaban en ambos oligonucleótidos.

Inicialmente se realizó la hibridación (*annealing*) de los oligonucleótidos. Para ello, se mezcló la cantidad de 1  $\mu\text{M}$  de cada oligonucleótido y 0,1 mg/mL de BSA en tampón CutSmart® (NEB) y un volumen final de 96,5  $\mu\text{L}$ . La mezcla se incubó en el termociclador con el siguiente ciclo:

- 2 minutos a 95 °C.
- 45 minutos de rampa de temperatura descendente hasta los 25 °C.
- Descenso de la temperatura hasta 4°C, donde se mantiene.

Tras la hibridación, se procedió con la extensión de los oligonucleótidos por sus extremos 3' para completar el fragmento de interés. Se añadió a la reacción 2,5  $\mu\text{L}$  de dNTPs, a una concentración de 10 mM cada uno en la mezcla de la que se tomaron, y 2 unidades de la polimerasa Klenow (Roche). La reacción de extensión se incubó 1 hora a 37 °C y, finalmente, la enzima se inactivó 20 minutos a 75 °C.

#### 6.1.10. Secuenciación de pequeños fragmentos de ADN

Las reacciones de secuenciación de ADN fue llevada a cabo por la empresa Secugen (Madrid), usando como molde productos de PCR o preparaciones de ADN plasmídico. Como cebadores se utilizaron oligonucleótidos universales o específicos del fragmento a secuenciar. Para la comparación de las secuencias con las bases de datos y el análisis de las pautas abiertas de lectura se utilizó el paquete de herramientas BLAST (Altschul y col., 1990) disponible en la página web del NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Para alinear secuencias se usó el programa online Clustal Omega (EMBL-EBI, Hinxton, UK) manteniendo los parámetros determinados por defecto.

#### 6.1.11. Construcción de plásmidos y de estirpes de TFA

##### Construcción de la estirpe mutante MPO642

Para la obtención del mutante de sustitución del gen *suhB* (SGRAN\_0363) por un gen de resistencia a kanamicina, se construyó el plásmido pMPO1150 basado en el plásmido pEX18Tc (Hoang y col., 1998), el cual no es replicativo en TFA. El plásmido pEX18Tc contiene el gen *sacB* cuya expresión resulta letal para las bacterias que se cultivan en presencia de sacarosa. Se construyeron los plásmidos pMPO1149 y pMPO1150 para la construcción de dicho mutante, los cuales se describen a continuación.

**pMPO1149:** Plásmido derivado del vector pBluescript II SK (+) conteniendo 726 pb de la región aguas arriba del pequeño ARN *SuhB* (Flanqueante R, comienza en la tercera base aguas arriba del +1 de *suhB* determinado por dRNA-seq) y 1054 pb de la región aguas abajo (Flanqueante L, comienza 86 pb aguas abajo del final de *suhB* determinado por dRNA-seq), teniendo en cuenta el transcrito definido por dRNA-seq). Las flanqueantes se obtuvieron por PCR sobre ADN genómico de TFA con los oligonucleótidos *suhB\_F2\_HD* y *suhB\_R2\_BH* y *suhB\_F1\_BH* y *suhB\_R1\_HD*, respectivamente. El producto de ambas PCRs se cortó con *NcoI* y *BamHI* y se clonaron juntas en pBluescript II SK (+) en la diana *BamHI*. A continuación, el plásmido se cortó con *NcoI* (diana situada entre ambas flanqueantes) y se romizó, donde se clonó el casete de resistencia a kanamicina aislado como un fragmento *HindIII* de pUTminiTn5 y romizado.

**pMPO1150:** Plásmido derivado del vector pEX18Tc en el que se clona, en la diana *BamHI*, el inserto de 1780 pb de pMPO1149 digerido con la misma enzima de restricción.

El plásmido pMPO1150 se conjugó en *S. granuli* TFA silvestre y se seleccionaron candidatos *Str<sup>r</sup>* *Tc<sup>r</sup>* y *Km<sup>r</sup>* tras un periodo de incubación de 5 a 7 días en placas de MML sin

## ***Materiales y Métodos***

sacarosa. Dichos clones, por su fenotipo, debían haber integrado el plásmido en el cromosoma por un evento de recombinación a través de cualquiera de las dos flanqueantes, lo cual se corroboró por PCR. Una vez comprobados los candidatos, se seleccionaron varios para incubarlos durante 3 días en MML líquido sin tetraciclina, tiempo tras el cual se sembraron en placas de MML conteniendo sacarosa al 10 %. Estas placas se cultivaron unos 5 días y se seleccionaron los candidatos Str<sup>r</sup> Km<sup>r</sup> y Sac<sup>r</sup>.

### **Construcción de la estirpe MPO501**

Para obtener la estirpe con un reemplazo alélico del gen *hfq* por *hfq::3xFlag* (MPO501) se construyeron los plásmidos pMPO1806 y pMPO1808.

**pMPO1806:** Inicialmente, el vector pGG1, que porta el epítipo 3xFlag, se amplificó por PCR con la pareja de oligonucleótidos divergentes JVO5142 y CSO-0073. A continuación, se amplificó por PCR el gen *hfq* (705 pb), utilizando como molde ADN genómico de la estirpe silvestre de TFA con los oligonucleótidos TFA\_Hfq\_up\_AatII, que lleva el sitio de corte de la enzima AatII, y TFA\_Hfq\_Flag, cuyo extremo 5' está fosforilado. Vector e inserto se digirieron con la enzima AatII y se ligaron, obteniéndose el plásmido pMPO1806.

**pMPO1808:** El plásmido pMPO1806 se amplificó por PCR con los oligonucleótidos divergentes pMPO1806\_NcoI, que lleva el sitio de corte de la enzima NcoI, y TFA\_Hfq\_Flag. La región aguas abajo de *hfq* (~700 pb) se amplificó utilizando los oligonucleótidos Hfq\_end\_NcoI, con un sitio NcoI, y GTPase\_down\_Xba, que porta una diana XbaI. Vector e inserto se digirieron con las enzimas de restricción NcoI y XbaI y se ligaron para obtener el plásmido pMPO1808.

Finalmente, el plásmido pMPO1808 se digirió con las enzimas AatII y XbaI para obtener el fragmento lineal del inserto que contiene, en este orden, el alelo *hfq::3xFlag*, un gen de resistencia a kanamicina y el gen codificante para una GTPasa que se encuentra aguas abajo del gen *hfq* en el genoma de TFA. Los fragmentos lineales se purificaron de la reacción enzimática con el kit GFX (GE healthcare) y se electroporaron en células electrocompetentes de la estirpe silvestre de TFA. Se seleccionaron en medio rico MML los candidatos Str<sup>r</sup> Km<sup>r</sup>.

### **Construcción de estirpes con fusiones génicas**

Para la obtención de las estirpes MPO641 y MPO644, las cuales portan la fusión transcripcional *P<sub>suhB</sub>::lacZ*, se generó el plásmido pMPO1151 cuya construcción se detalla a continuación.



**pMPO1151:** Plásmido derivado de pIC552 que porta una fusión transcripcional del promotor de *suhB* al gen *lacZ*. El plásmido contiene 726 pb de la región aguas arriba del gen *suhB* (Flanqueante R, comienza en la tercera base aguas arriba del +1 determinado por dRNA-seq). Dicho inserto se obtuvo por PCR sobre ADN genómico de TFA silvestre con los oligonucleótidos *suhB\_F2\_HD* y *suhB\_R2\_BH*. El producto de PCR, sin cortar, se clonó in situ en pIC552 cortado con *SmaI* (enzima que deja extremos romos) seleccionándose la orientación adecuada para permitir la expresión del gen *lacZ* bajo el promotor de *suhB*.

El vector pMPO1151 se introdujo por electroporación en la estirpe silvestre de TFA, donde se integró en el cromosoma. Los candidatos *Str<sup>r</sup> Ap<sup>r</sup>* comprobados se denominaron estirpe MPO641 (*PsuhB::lacZ*). Del mismo modo, pMPO1151 se introdujo en células electrocompetentes de MPO209 (estirpe de TFA con una inserción de un miniTn5 en el gen *phaC*) y se seleccionaron los candidatos *Str<sup>r</sup> Ap<sup>r</sup>*, obteniéndose la estirpe MPO644 (mutante de inserción en *phaC*; *PsuhB::lacZ*).

Para la obtención de la estirpe MPO643 ( $\Delta$ *suhB::kan*; *thnC::lacZ*), que contiene la fusión traduccional *thnC::lacZ*, se electroporó el vector pIZ1002 (Martínez-Pérez y col., 2004) en células electrocompetentes de la estirpe MPO642 y se seleccionaron candidatos en medio rico MML que fuesen *Str<sup>r</sup> Km<sup>r</sup> Ap<sup>r</sup>*.

Para la complementación funcional de la estirpe mutante MPO643 se obtuvo el plásmido pMPO1152 cuya construcción se detalla a continuación.

**pMPO1152:** Plásmido derivado de pLAFR3 que contiene clonado un fragmento de 235 pb con la secuencia del gen *suhB*, hasta la última T de su secuencia, y su región promotora, hasta 155 pb aguas arriba del TSS predicho. Dicho fragmento se amplificó con la pareja de oligonucleótidos *suhB\_compR*, que contiene una diana *BamHI*, y *suhB\_compF*, el cual porta una diana *HindIII*. El fragmento se digirió con las enzimas *BamHI* y *HindIII* y se clonó en pLAFR3 digerido con las mismas enzimas. El promotor de *suhB* queda en la misma dirección que el promotor *Plac* de pLAFR3 para evitar una transcripción opuesta que hibride con *SuhB*.

## 6.2. Manipulación de ARN

### 6.2.1. Aislamiento de ARN de *S. granuli* TFA

Para obtener el ARN total de células de TFA, los cultivos se enfriaron en hielo y la biomasa se recogió centrifugando a 9790 *g* y 4 °C. La biomasa fue obtenida de 10 mL de cultivo en fase exponencial.

El sedimento de células se resuspendió en 1,5 mL de Tripure Isolation Reagent (Tri Reagent LS, Molecular Research Centre) y la mezcla se incubó 10 minutos a 60 °C, mezclando por inversión, para lisar las células. Posteriormente, la mezcla lisada se centrifugó a 16.100 g durante 10 minutos a 4 °C para sedimentar los restos celulares. El sobrenadante se transfirió a un tubo Phase Lock Gel™ (5PRIME) precompactado, al que se añadieron 200 µL de cloroformo. En este punto, la muestra se homogeneizó y se mantuvo en hielo durante 15 minutos, tiempo tras el cual se centrifugó a 16.100 g durante 15 minutos y a temperatura ambiente. Para la extracción de la fase acuosa, se añadieron al tubo 400 µL de fenol:cloroformo:isoamílico (25:24:1) y se centrifugó de nuevo a 16.100 g durante 5 minutos y a 4 °C. La fase acuosa se transfirió a un nuevo tubo al que se añadieron 750 µL de isopropanol y se incubó 10 minutos en hielo para volver a centrifugar a 16.100 g durante 10 minutos y a 4 °C. Se eliminó el sobrenadante y el precipitado se lavó con etanol al 70 % (v/v). Finalmente la muestra se secó en la campana de flujo laminar y se resuspendió en 150 µL de agua fría tratada con dietilpirocarbonato, compuesto que inactiva las RNasas (agua DEPC).

Para eliminar el ADN que pudiese permanecer en la muestra de ARN obtenida, esta se trató con DNasa I, utilizando para ello el kit DNA free (Ambion) y siguiendo las instrucciones del fabricante. Además, para descartar la presencia de ADN contaminante se realizó una PCR con los oligonucleótidos f27 y r519 (Hugenholtz et al., 1998) que amplifican el gen codificante para el ARN ribosómico 16S de bacterias y se comprobó a través de electroforesis en gel de agarosa la ausencia de producto amplificado. Finalmente, las muestras se purificaron usando el kit RNeasy mini kit (Quiagen) y se eluyeron en 50 µL de agua DEPC.

### **6.2.2. Síntesis de transcritos de ARN mediante transcripción *in vitro***

Los transcritos utilizados en los ensayos de retardo de movilidad en gel para detectar interacción entre ARNs se obtuvieron mediante transcripción *in vitro* (TIV).

Las secuencias moldes para la transcripción de *suhB*, 5'UTR-*thnR* y 5'UTR-*thnR* con los sitios de interacción con SuhB mutados conteniendo el promotor de T7 se obtuvieron por PCR con las parejas de oligonucleótidos (*suhB\_IVT\_T7\_F*, *suhB\_IVT\_R*), (*thnR\_IVT\_T7\_F*, *thnR\_IVT\_R*) y (*5thnR\_mut\_fwd*, *5thnR\_mut\_rev*) respectivamente. Las dos primeras secuencias molde se amplificaron mediante una PCR convencional, mientras que la versión mutada de 5'UTR-*thnR* se sintetizó por extensión con la polimerasa Klenow (Roche) de los oligonucleótidos mencionados anteriormente, los cuales solapaban entre sí parcialmente. Posteriormente, los fragmentos obtenidos por PRC se clonaron en los sitios EcoRI y HindIII del vector pUC19 (Yanisch-Perron y col., 1985), obteniéndose los plásmidos pMPO1155 (*suhB*), pMPO1156 (5'UTR-*thnR*) y pMPO1157 (5'UTR-*thnR* mutado).

El ADN molde para las reacciones de TIV se obtuvo digiriendo los plásmidos anteriores con HindIII y NdeI y recuperando el fragmento lineal que contenía cada secuencia bajo el control del promotor T7. De este modo, el final de la transcripción tendría lugar en el extremo 3' del fragmento lineal. Posteriormente, 1 µg del fragmento purificado se utilizó como molde en la reacción de TIV que contenía, en un volumen final de 20 µL en el tampón suministrado por la casa comercial, 40 U de la ARN polimerasa de T7 (20 U/µL, Roche), 20 U de RNaseOUT™ (40 U/µL, Invitrogen) y 1 mM de cada uno de los NTPs (ATP, GTP, CTP y UTP).

Las reacciones de TIV se incubaron entre 6 y 8 horas a 37 °C, tiempo tras el cual se pararon con la adición de 4 µL de EDTA 100 mM (pH 8). Para eliminar el ADN molde, se incorporaron a las reacciones 20 U de DNase I RNase-free (10 U/µL, Roche) y se mantuvieron 15 minutos más a 37 °C. A continuación, se incrementó el volumen de reacción a 100 µL con agua DEPC y las muestras se fenolizaron añadiendo 1 volumen (100 µL) de fenol:cloroformo:isoamílico (25:24:1) en tubos de Phase Lock Gel™ (5PRIME) precompactados. Dichos tubos se centrifugaron 12 minutos a 13000 rpm y la fase acuosa (~100 µL) se pasó por una columna MicroSpin™ G-25 (Amersham Pharmacia Biotech). Finalmente, el ARN se precipitó unas 14 horas a -20 °C con 0,1 volúmenes de acetato de sodio 3M (pH 5,2) y 2,5 volúmenes de etanol al 96 %.

Para recuperar el ARN precipitado, las muestras se centrifugaron a 4 °C durante 30 minutos a 13000 rpm. Luego, se eliminó el sobrenadante y se realizó un lavado con 250 µL de etanol al 70 %, centrifugando a 4 °C durante 10 minutos a 13000 rpm y eliminando el sobrenadante. El ARN precipitado se dejó secar a temperatura ambiente en la campana de flujo laminar y, finalmente, se resuspendió en 33 µL de agua DEPC durante 5 minutos a 65 °C.

### **6.2.3. Electroforesis de ARN**

Para visualizar y comprobar la integridad de las muestras de ARN total, estas se visualizaron en geles de agarosa al 1 % (p/v), del mismo modo descrito para la electroforesis de ADN. En el caso de los transcritos obtenidos por TIV y tras su posterior manipulación, la visualización y comprobación de la integridad del ARN se realizó mediante electroforesis vertical en geles desnaturizantes de poliacrilamida 29:1 10 % (p/v)-urea 8 M en tampón de carrera TBE 1X y a un amperaje constante de 4 mA.

Para los geles que se utilizaron en los *Northern Blots* (ver siguiente sección), se cargaron 5 µg de ARN total en una gel desnaturizante de poliacrilamida 19:1 al 12 % (p/v)-urea 8,3 M, en tampón de carrera TBE 1X. Como marcador de peso molecular se utilizaron diluciones

apropiadas del marcador pUC mix marker 8 (Fermentas) marcado radiactivamente con  $P^{32}$  (dependiendo de la antigüedad del marcador) y se cargaron 5  $\mu$ L en uno de los pocillos del gel. Posteriormente, el gel se corrió a 150 V durante 2 h mediante electroforesis vertical. En todos los casos se utilizó material y soluciones libres de ARNasas y tratados con agua DEPC.

TBE 1X: Tris-ácido bórico 90 mM, EDTA 2 mM (pH 8,3)

### **6.2.4. Detección de secuencias de ARN por hibridación en membrana (*Northern Blot*)**

La transferencia del ARN desde el gel a una membrana Hybond-N+ (GE Healthcare) se realizó mediante la aplicación de una corriente eléctrica de 50 V durante 1 hora a 4 °C en el Sistema Trans-Blot® SD semi-dry transfer cell (BioRad) según las especificaciones del fabricante. Por último, se fijó el ARN a la membrana mediante la aplicación de luz ultravioleta. Una vez en este punto, la membrana se puede almacenar en papel de filtro o hibridar directamente.

La sonda para la hibridación consistió en un oligonucleótido de ADN de entre 18 y 22 pb complementario al ARN a detectar. Este oligonucleótido se marcó radiactivamente en el extremo 5' con  $\gamma[^{32}P]$ -dATP mediante una reacción de fosforilación catalizada por la enzima polinucleótido quinasa (PNK) de T4 (Roche). Dicha reacción se realizó en un volumen final de 20  $\mu$ L, mezclando 10 pmoles de oligonucleótidos (1  $\mu$ L de oligonucleótidos a 10  $\mu$ M), 30  $\mu$ Ci de  $\gamma[^{32}P]$ -dATP, 10 U de PNK y el tampón que proporciona la casa comercial. La mezcla se incubó durante 1 hora a 37 °C y seguidamente la PNK se inactivó a 95 °C durante 5 minutos. Para eliminar los nucleótidos y la marca no incorporada se pasó todo el volumen de reacción por una columna de sephadex G25 (GE Healthcare). Finalmente se cuantificó la marca con un contador de centelleo. En este punto, la sonda puede almacenarse a -20 °C.

La hibridación de la membrana con la sonda marcada se realizó en un horno de hibridación Hybridizer HB-1000 (UVP LLC). Inicialmente, la membrana se prehibridó un mínimo de 1 h a 42 °C con solución de prehibridación precalentada 10 minutos a 65 °C. La hibridación posterior se llevó a cabo durante unas 14-16 horas a 42 °C en solución de prehibridación a la que se añadieron los 20  $\mu$ L de sonda marcada y purificada. Posteriormente, la membrana se lavó a 30 °C tres veces durante 20, 15 y 15 minutos con SSC y 0,1 % SDS, estando el SSC a una concentración de 5X, 1X y 0,5X respectivamente.

Tras los lavados, la membrana se expuso durante al menos 12 horas en una pantalla sensible a radiaciones gamma (Storage Phosphor Screen, GE Healthcare). Finalmente, la

pantalla se escaneó con un escáner Typhoon 9410 (GE Healthcare) para visualizar la señal emitida.

### 6.3. Cuantificación de ácidos nucleicos

Para cuantificar las distintas preparaciones de ADN de doble cadena (ADN plasmídico, fragmentos de ADN purificados, ADN total, etc.), se midió la absorbancia a 260 nm en un espectrofotómetro Nanodrop ND-1000 Spectrophotometer (Thermo Scientific), aplicando un coeficiente de extinción de  $50 \text{ ng cm } \mu\text{L}^{-1}$  para calcular la concentración de la muestra. En el caso de las preparaciones de ARN total y de transcritos procedentes de transcripción *in vitro*, la cuantificación se realizó del mismo modo pero aplicando un coeficiente de extinción de  $40 \text{ ng cm } \mu\text{L}^{-1}$ .

## 7. Detección de proteínas mediante inmunodetección (*Western Blot*)

### Preparación de extractos celulares

Como paso previo a la detección de proteínas concretas por *Western Blot*, se procedió a la obtención de extractos de células enteras. Dichos extractos se obtuvieron a partir de células de *S. granuli* cultivadas en distintos medios, partiendo de un volumen de 8 mL y recogiendo la biomasa por centrifugación. Para ello, se normalizaron las muestras en función de su D.O.<sub>600</sub> de forma que por cada unidad de D.O.<sub>600</sub> se añadieron 10,9  $\mu\text{L}$  de agua bidestilada y el mismo volumen de tampón de carga 2X con  $\beta$ -mercaptoetanol 10 % (v/v). Las muestras se hirvieron durante 10 minutos y se centrifugaron a velocidad máxima para sedimentar los restos celulares.

Tampón de carga 2x: Tris-HCl 160 mM pH 8, glicerol 20 % (v/v), SDS 4 % (p/v), azul de bromofenol 0,1 % (p/v).

### Electroforesis SDS-PAGE

Para separar las proteínas de una preparación de extracto crudo, las muestra se sometieron a electroforesis en gel de poliacrilamida preparado con las soluciones del kit 12 % TGX Stain-Free™ FastCast™ Acrylamide gels (BioRad) según las indicaciones del fabricante y con un grosor de 1 mm. Con esta tecnología, la visualización del resultado de la electroforesis no requirió de tinción puesto que las proteínas se modificaron en sus residuos de triptófano adquiriendo la capacidad de emitir fluorescencia.

Una vez polimerizado el gel, se cargaron 10  $\mu\text{L}$  de las muestras de extracto crudo y, como control, se añadieron 3 ng de ThnR purificada (5  $\mu\text{L}$  de la proteína a una concentración 19 nM). La electroforesis se llevó a cabo en un sistema vertical Mini Protean® III (BioRad) con

## **Materiales y Métodos**

Tank Buffer 1X a una intensidad de corriente fija de 30 mA por gel y durante aproximadamente 45 minutos. Posteriormente, los geles se lavaron con agua bidestilada y se activaron y visualizaron a través del sistema de imagen ChemiDoc™ (BioRad).

Tank buffer 1X: Trizma base 25 mM, glicina 19,2 mM y SDS 0,1 % (p/v).

### **Transferencia a membrana, inmunodetección y revelado**

Las proteínas, separadas por SDS-PAGE, se transfirieron a una membrana de nitrocelulosa usando un sistema de transferencia semi-seca Trans-Blot® Turbo™ Transfer System (BioRad). La membrana y dos pilas de papel de filtro (*stacks*, BioRad) de transferencia se activaron durante 3 minutos en tampón de transferencia 1X. A continuación, se montó el sistema de transferencia colocando los elementos, desde el ánodo al cátodo, en el siguiente orden: una de las pilas de papel de filtro, la membrana, el gel y la segunda tanda de papel de filtro. La transferencia de las proteínas desde el gel hasta la membrana se realizó a 25V durante 7 minutos seleccionando la opción Turbo del sistema. Para evaluar la eficiencia de la transferencia, las proteínas (activadas previamente mediante la tecnología de *Stain-Free*) se visualizaron a través del sistema de imagen ChemiDoc™ (BioRad).

Una vez comprobado que la transferencia se produjo correctamente, las membranas se lavaron 3 veces durante 5 minutos en tampón TTBS 1X y se incubaron con solución de bloqueo, al menos 1 hora, a temperatura ambiente y en agitación. Posteriormente, se añadió el anticuerpo primario limpio y diluido 1:1000 en solución de bloqueo, y se incubó toda la noche a 4 °C en agitación suave. La limpieza del anticuerpo se llevó a cabo mediante la incubación del mismo con sobrenadante del hervido de extracto bacteriano carente de la proteína a detectar. En esta tesis, se ha empleado únicamente el anticuerpo primario policlonal anti-ThnR proporcionado por la casa comercial Mologen Molecular Medicines. Tras la incubación con el anticuerpo primario, se realizaron 4 lavados de 15 minutos con tampón TTBS 1X y, a continuación, se incubó 2 horas con una dilución 1:5000 del anticuerpo secundario anti-rabbit peroxidase (Sigma-Aldrich) preparada en nueva solución de bloqueo. La peroxidasa conjugada en dicho anticuerpo genera quimioluminiscencia en presencia de un sustrato adecuado. Luego, se volvió a lavar la membrana 4 veces durante 15 minutos con TTBS 1X.

Para el revelado de la membrana, esta se incubó durante 5 minutos con una mezcla 1:1 (v:v) de los sustratos del kit Immun-Star™ AP Chemiluminescence (BioRad). Seguidamente, se reveló la señal en el sistema de imagen ChemiDoc™ (BioRad), a través del programa ImageLab (BioRad), con el cual también se llevó a cabo la posterior normalización y cuantificación de la señal, como se describe en (Gürtler y col., 2013) y según la guía *Western Blot Normalization Using Image Lab™ Software* de BioRad.

Tampón de transferencia 1X: 200 mL de Transfer Buffer 5X (BioRad), 600 mL de agua bidestilada y 200 mL de etanol grado biología molecular (~95-98 % de pureza).

TBS 5x: 75 mL Tris 1M, 800 mL NaCl 2,5 M, 50 mL H<sub>2</sub>O bidestilada.

TTBS 1X: Tween 20 al 0,05 % en TBS 1X.

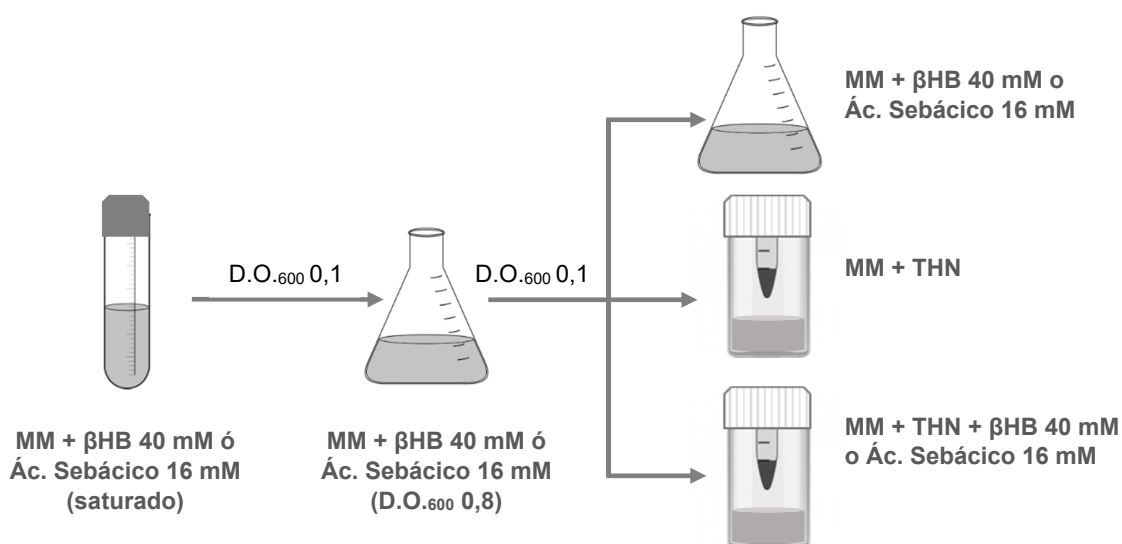
Solución de bloqueo: 5 % (p/v) de leche en polvo en TTBS 1X.

## 8. Análisis de la expresión génica

### 8.1. Ensayos de actividad $\beta$ -galactosidasa

#### 8.1.1. Obtención de células para la medida de expresión de fusiones a *lacZ*

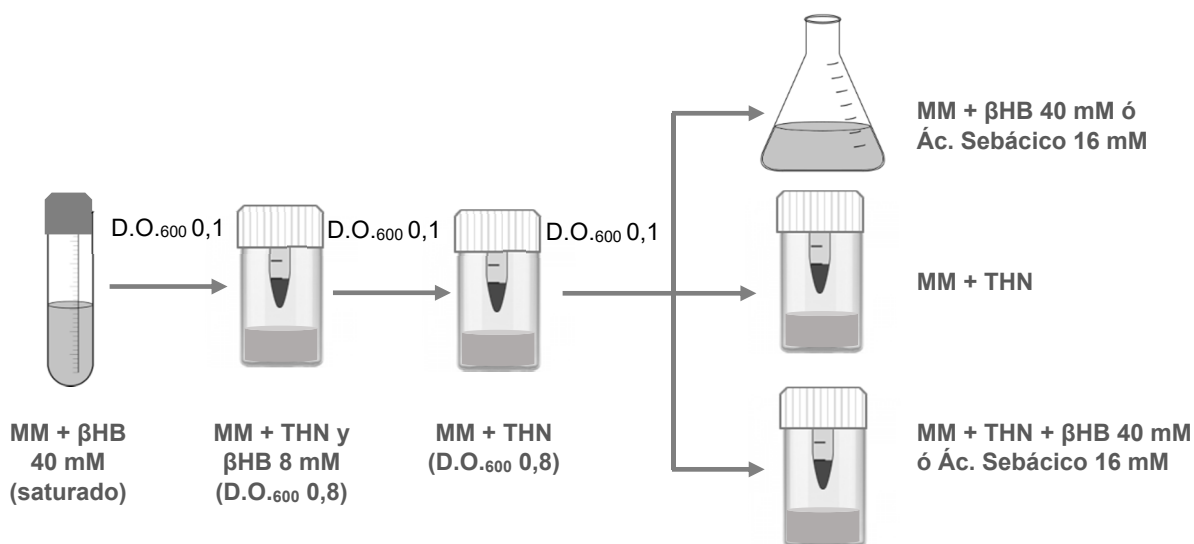
Para la medida de la expresión de los genes *thn* en *S. granuli* se utilizaron estirpes que llevaban la fusión traduccional *thnC::lacZ* insertada en el cromosoma. Dichas estirpes se cultivaron en condiciones de inducción (MM con tetralina), condiciones de represión catabólica (MM con tetralina y  $\beta$ -HB 40 mM o ácido sebácico 16 mM) y condiciones de no inducción (MM con  $\beta$ HB 40 mM o ácido sebácico 16 mM). En todos los casos los cultivos se obtuvieron de pre-inóculos cultivados hasta la saturación en  $\beta$ HB 40 mM o ácido sebácico 16 mM. Estos pre-inóculos se utilizaron para inocular nuevos cultivos intermedios, como se muestra en la Figura 16, en el mismo medio con el fin de garantizar que los cultivos finales fueran inoculados a partir de los intermedios en fase exponencial, aproximadamente a D.O.<sub>600</sub> de 0,8.



**Figura 16.** Obtención de cultivos para medida de la actividad  $\beta$ -galactosidasa de la fusión traduccional *thnC::lacZ*.

## Materiales y Métodos

En el caso de la medida de la expresión de *suhB* se emplearon cepas de *S. granuli* que portaban la fusión transcripcional *PsuhB::lacZ* integrada en el cromosoma. El procedimiento por el cual se obtuvieron las células para el posterior ensayo de actividad  $\beta$ -galactosidasa fue el mismo que en el caso de las cepas con la fusión *thnC::lacZ*, si bien, el pre-inóculo inicial y los preinóculos intermedios fueron distintos, como se muestra en la Figura 17, con el objetivo de inocular los medios finales a partir de un preinóculo crecido en MM con THN, donde la expresión de *suhB* es baja.



**Figura 17. Obtención de cultivos para medida de la actividad  $\beta$ -galactosidasa de la fusión transcripcional *PsuhB::lacZ*.**

En todos los casos, para inocular de un medio al siguiente, se centrifugó la cantidad conveniente de cultivo para que la D.O.<sub>600</sub> del nuevo fuese de 0,1. La centrifugación se realizó durante 5 minutos a velocidad máxima en una centrífuga de mesa y se eliminó por completo el sobrenadante. El sedimento de células se resuspendió en tampón fosfato y se inoculó en el nuevo medio de cultivo. Por último, para la realización del ensayo de actividad  $\beta$ -galactosidasa, se recogieron células de los distintos cultivos a diferentes tiempos desde su inoculación.

### 8.1.2. Medida de la actividad $\beta$ -galactosidasa

Los ensayos de actividad  $\beta$ -galactosidasa se realizaron según el protocolo descrito por Miller (Miller, 1972) con algunas modificaciones. Un volumen de 100  $\mu$ L de una dilución apropiada del cultivo se añadió a una mezcla de 650  $\mu$ L de tampón Z, 20  $\mu$ L de SDS 0,1 % (p/v) y 30  $\mu$ L de cloroformo. La mezcla formada se agitó vigorosamente con el vórtex para permeabilizar las células y se incubó 10 minutos a 30 °C en un baño. Seguidamente, se añadieron 200  $\mu$ L de ONPG (4 mg/L en tampón Z) a cada muestra y se incubaron de nuevo a



30 °C hasta la aparición de coloración amarilla. Tras el cambio de color, la reacción se paró mediante la adicción de 500 µL de Na<sub>2</sub>CO<sub>3</sub> 1M, homogeneizando con el vórtex y depositando la muestra en hielo. Las muestras se centrifugaron a temperatura ambiente durante 15 minutos y a 9300 g para separar el cloroformo de los restos celulares. Por último, se midió la absorbancia a 420 nm de la fase acuosa. La actividad se calculó aplicando la siguiente fórmula:

$$\text{Actividad } \beta \text{ galactosidasa (Unidades Miller)} = \frac{A_{420}}{V \cdot D.O._{600} \cdot t} \cdot 1000$$

Donde A<sub>420</sub> corresponde a la absorbancia de la reacción a 420 nm de longitud de onda, V al volumen de cultivo utilizado en la reacción (0,1 mL), D.O.<sub>600</sub> a la densidad óptica del cultivo a 600 nm de longitud de onda y t al tiempo de reacción en minutos.

## 8.2. Experimentos de transcriptómica (dRNA-seq)

La expresión global de los genes de TFA se analizó a través de la secuenciación el ARN extraído de células recogidas en fase exponencial de crecimiento (D.O.<sub>600</sub> ~ 0,8), tanto en medio mínimo con tetralina como en medio mínimo con βHB 40 mM. La tecnología empleada, conocida como RNA-seq diferencial (dRNA-seq del inglés *differential RNA-seq*) permitió además la identificación global de posibles inicios de la transcripción (TSS del inglés *Transcriptional Start Site*). El experimento se realizó en colaboración con los doctores Cynthia M. Sharma y Jorg Vogel en el Research Center for Infectious Diseases (ZINF) de la Universidad de Würzburg (Alemania), siguiendo el protocolo descrito en Sharma y col. (2010) (material suplementario) y Bischler y col. (2015).

Brevemente, en la técnica de dRNA-seq, la muestra de ARN a analizar se dividió en dos, una se trató con la enzima TEX (*Terminator 5'-Phosphate-dependent Exonuclease*) mientras que la otra no. Esta exonucleasa digiere específicamente los transcritos con un único fosfato en su extremo 5', los cuales han sido procesados desde un transcrito primario. De esta forma, la muestra se enriquece en transcritos primarios. A continuación, ambas muestras de ARN, tratada y sin tratar con TEX, se utilizan para sintetizar librerías de ADNc y secuenciarlas. El número de secuencias obtenidas de cada muestra se normalizan para su comparación y visualización con la herramienta IGB (*Integrative Genome Browser*) (Freese y col., 2016).

La detección de los TTS se realizó a través del programa TSSpredator (Dugar y col., 2013). Como datos de entrada se utilizaron los ficheros con formato WIG procedentes de la normalización de los datos de dRNA-seq correspondientes a las muestras tratadas con TEX y sin tratar, en cada una de las condiciones estudiadas.

## **9. Detección y validación de la interacción ARN-ARN**

### **9.1. Detección *in silico* de interacciones ARN-ARN**

Las dianas de pequeños ARNs se detectaron utilizando el programa IntaRNA (Busch y col., 2008). Como entrada a la herramienta se utilizó la secuencia del pequeño ARN regulador y un fichero FASTA con las secuencias de ARN donde se querían localizar las dianas. En el caso de la detección de la interacción en las regiones 5' no traducidas, se extrajeron del genoma de TFA las secuencias comprendidas entre las 150 pb aguas arriba y 150 pb aguas abajo del codón de inicio de cada gen, mediante un script codificado en lenguaje Perl (Anexo IV). En dicho programa también se sustituyeron las Ts por Us.

### **9.2. Sistema reportero basado en fusiones a GFP**

Para validar la interacción entre el pequeño ARN SuhB y la región 5' UTR del ARN mensajero de *thnR* se realizaron ensayos de interacción basados en un sistema modular que utiliza la proteína verde fluorescente (GFP) como indicador para el estudio de la regulación post-transcripcional. El experimento se realizó *in vivo* en células de TOP10F' según el protocolo descrito por Urban y Vogel (2007) y Corcoran y col. (2012).

#### **9.2.1. Construcción de vectores**

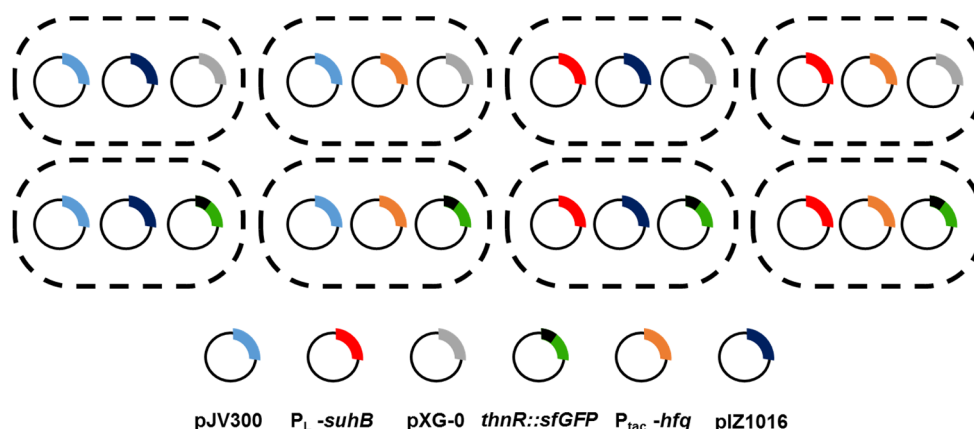
El pequeño ARN se clonó bajo el control del promotor P<sub>L</sub>lacO1 en el plásmido pZE12-luc. Para ello, dicho vector se amplificó con los oligonucleótidos PllacoD y PllacoB-1, obteniendo un fragmento de ~3,9 kb que se digirió con la enzima XbaI. Posteriormente, se cortó de gel y purificó la banda de ~2,2 kb que contenía el promotor P<sub>L</sub>lacO1 desde la posición -1. Por otra parte, el pequeño ARN se amplificó con los oligonucleótidos SuhB\_F\_phosp (fosforilado en el 5') y SuhB\_R\_XbaI para obtener la secuencia desde el +1 de la transcripción hasta 108 nucleótidos aguas abajo de la terminación del transcrito de *suhB*. Este fragmento, digerido con XbaI, se clonó en el fragmento de pZE12-luc obtenido anteriormente, obteniendo el plásmido pMPO1153 (P<sub>L</sub>-*suhB*).

Por otra parte, la región desde el +1 hasta los 27 primeros codones de *thnR*, se fusionó traduccionalmente a una proteína verde fluorescente super-plegada, sfGFP, *superfolder Green Fluorescence Protein* (Pédelacq y col., 2006) en el plásmido pXG10-SF, obteniendo el vector pMPO1154 (*thnR::sfGFP*). Para ello, el fragmento del gen *thnR*, amplificado con los oligonucleótidos thnR\_F\_NsiI y thnR\_R\_NheI utilizando como molde ADN genómico de TFA, se clonó de forma dirigida en los sitios NsiI y NheI del vector pXG10-SF.

La región de *thnR* con los dos posibles inicios de traducción TTG sustituidos por ATG se consiguió por PCR solapante con los oligonucleótidos mutagénicos *thnR\_solap\_ATGs\_F* y *thnR\_solap\_ATGs\_R* y los externos *pXG10\_ups\_gfp\_F* y *sfGFP\_pXG10\_out\_R*, utilizando como molde el vector intermedio descrito en el párrafo anterior. Los posibles inicios de la traducción correspondían al descrito en López-Sánchez y col. (2009) y al predicho por la herramienta Prodigal en el proceso de anotación. Sólo cuando ambos TTG se sustituyeron por ATG se consiguió una fusión que emitía fluorescencia.

### 9.2.2. Ensayos de fluorescencia

Como paso previo a los ensayos de fluorescencia, la cepa *E. coli* TOP10F' se transformó por el método de TSS (ver apartado 5.1.1 método 2 de materiales y métodos) con los distintos plásmidos, de forma secuencial, para obtener las combinaciones de plásmidos que se muestran en la Figura 18.



**Figura 18.** Representación de las distintas combinaciones de plásmidos para el ensayo de interacción entre 5'UTR-*thnR* y *SuhB*.

Para comenzar el ensayo de interacción, se inocularon células de TOP10F' llevando cada una de las combinaciones de plásmidos anteriores en 3 mL de LB con los antibióticos correspondientes y se cultivaron 14 horas a 37 °C. Posteriormente, se diluyeron los inóculos 100 veces en LB y por duplicado, adicionando a una de las réplicas 1 mM de IPTG para inducir la expresión de *suhB* y *hfq* en las células transformadas con pMPO1153 ( $P_L$ -*suhB*) y/o pMPO1819 ( $P_{tac}$ -*hfq*).

A continuación, todos los cultivos se incubaron 6 horas a 37 °C, momento tras el cual se diluyeron 2 veces en 1 mL de PBS 1X. Para ello, 500  $\mu$ L del cultivo se centrifugaron durante 3 minutos a 13000 rpm, se retiró el sobrenadante, se añadió 1 mL de PBS 1X y se realizaron dos lavados más con PBS 1X del mismo modo. Tras los lavados, se resuspendió el sedimento

## **Materiales y Métodos**

celular en 1 mL de PBS 1X, quedando la muestra lista para el análisis. Finalmente, se colocaron 200  $\mu$ L de las muestras anteriores en una placa multipocillo para medir la densidad óptica a 600 nm y la fluorescencia emitida por la proteína GFP (excitación a 435 nm y emisión a 535 nm) en un fluorímetro (Tecan). Como blanco, se utilizó PBS 1X, incorporando 200  $\mu$ L en todos los pocillos que rodeaban las muestras.

Los valores obtenidos en la medida del fluorímetro se trataron de la siguiente forma: (1) A cada valor de D.O.<sub>600</sub> y fluorescencia se le restó la media de los blancos. (2) Cada valor de fluorescencia se dividió por la D.O.<sub>600</sub> correspondiente. (3) La fluorescencia específica de la fusión *thnR::sfGFP* (pMPO1154) en cada caso se calculó como [fluorescencia(pMPO1154(*thnR::sfGFP*) + sRNA + Z) – fluorescencia(pXG0 + sRNA + Z)], donde sRNA hace referencia a pMPO1153 (SuhB) o pJV300 (pequeño ARN control) y Z a pMPO1819 (*hfq*) o pIZ1016 (plásmido vacío). El cálculo se realizó de la misma forma para los cultivos no inducidos como para los inducidos con IPTG.

PBS 10X: NaCl 80 g/L, KCl 2,01 g/L, Na<sub>2</sub>HPO<sub>4</sub> 14,2 g/L, KH<sub>2</sub>PO<sub>4</sub> 2,72 g/L. Ajustar a pH 7,4. Autoclavar.

### **9.3. Ensayo de retardo de la movilidad en gel (EMSA, ARN-ARN)**

Para comprobar si un pequeño ARN y su posible diana interaccionaban entre sí, se realizaron EMSAs empleando los transcritos de ambos elementos, los cuales se obtuvieron por transcripción *in vitro* (ver apartado 6.2.2 de materiales y métodos).

#### **Marcaje radiactivo de transcritos de ARN**

Los transcritos de ARN se marcaron radiactivamente en su extremo 5' con  $\gamma$ [<sup>32</sup>P]-dATP. Para ello, 25 pmol de ARN se defosforilaron con 1 U de fosfatasa alcalina (SAP, *Shrimp Alkaline Phosphatase*, Affymetrix) en tampón H de Roche y un volumen final de 20  $\mu$ L. La reacción se incubó a 37 °C durante 90 minutos. Posteriormente, se incrementó el volumen de las reacciones a 100  $\mu$ L con agua DEPC y se fenolizaron como se describe en el apartado 6.2.2 de materiales y métodos para finalmente resuspenderlas en 17  $\mu$ L de agua DEPC.

Para el marcaje del extremo 5', 15  $\mu$ L (~22 pmoles) de ARN defosforilado, se incubaron con 2 U de la enzima polinucleótido quinasa (PNK) de T4 (Roche) en el tampón proporcionado por la casa comercial y 30  $\mu$ Ci de  $\gamma$ [<sup>32</sup>P]-dATP. La reacción se mantuvo a 37 °C durante 1 h y el ARN se purificó siguiendo el mismo procedimiento que en el apartado 6.2.2 de materiales y métodos. Una vez secado el precipitado, se resuspendió en un volumen adecuado de agua DEPC y se cuantificó la cantidad de radiactividad recuperada en un contador de centelleo. Se

asume una pérdida del 10 % de ARN con respecto al inicial en el momento del marcaje (Bak y col., 2015).

#### Ensayos de retardo de movilidad en gel de poliacrilamida

Las reacciones de unión ARN-ARN se realizaron en un volumen final de 20 µL en el tampón de unión, con 5 nM del transcrito de la diana marcado radiactivamente (sonda) y concentraciones crecientes del transcrito del pequeño ARN sin marcar. Como control, se incorporaron reacciones idénticas a las anteriores pero utilizando como sonda 5 nM del transcrito de la diana con los sitios de unión modificados. Las muestras se incubaron a 37 °C durante 20 minutos, parando la reacción con 4 µL de tampón de carga 6X.

Las mezclas de reacción ARN-ARN se resolvieron mediante electroforesis en gel nativo de acrilamida de 1,5 mm al 5 % (p/v) en tampón TBE 1X a 4 °C. Posteriormente, los geles se secaron sobre papel Whatman en un secador Hoefer™ Slab Gel Dryer CD 2000 (GE Healthcare) y se expusieron a una pantalla sensible a radiaciones gamma (Storage Phosphor Screen, GE Healthcare) durante 14 horas. Por último, las señales radiactivas se detectaron y visualizaron a través de un escáner Typhoon 9410 y se analizaron mediante el software ImageQuant (GE Healthcare).

Tampón de unión 2X: Tris-HCl 20 mM, pH8, NaCl 100 mM, KCl 100 mM, MgCl<sub>2</sub> 20 mM

### **10. Co-inmunoprecipitación de ARNs con Hfq**

El experimento se realizó en colaboración con los doctores Cynthia M. Sharma y Jorg Vogel en el *Research Center for Infectious Diseases* (ZINF) de la Universidad de Würzburg (Alemania), siguiendo el protocolo descrito en Sittka y col., 2008.

Brevemente, células de TFA silvestre y MPO501 (*hfq::3xFlag*) fueron recogidas en fase exponencial. Posteriormente, se obtuvieron extractos celulares de ambas muestras y se precipitaron con un anticuerpo anti-Flag. El ARN precipitado se purificó y se construyeron librerías de ADNc que se secuenciaron en el *Research Center for Infectious Diseases* (ZINF). Tras la normalización de las lecturas obtenidas, la visualización del resultado se realizó con el programa IGB.



## **RESULTADOS**





## 1. Análisis del genoma de *S. granuli* TFA

### 1.1. Secuenciación y ensamblado

El genoma de TFA se secuenció por la empresa LifeSequencing, S.L. utilizando la tecnología de pirosecuenciación 454 GS-FLX Titanium de Roche. Las pequeñas lecturas obtenidas con esta plataforma de secuenciación se almacenaron en un fichero con formato FASTA y en otro con formato SFF. Estas lecturas se ensamblaron *de novo* mediante los programas de ensamblado Celera Assembler (Myers y col., 2000) y Newbler (454 Life Sciences), obteniendo un total de 42 fragmentos de mayor tamaño con el primero y 53 con el segundo. Según los datos obtenidos por ambos programas (Tabla 8), se determinó que el ensamblador Celera Assembler producía un mejor resultado para continuar con el procedimiento de ensamblado y cierre de los huecos entre los fragmentos.

**Tabla 8. Características más relevantes del ensamblado conseguido con cada uno de los ensambladores automáticos.**

Característica	Celera Assembler	Newbler
Total de fragmentos	42	53
Total de bases (pb)	4649891	4654674
Longitud media de los fragmentos (pb)	110712	87799
Longitud del fragmento más pequeño (pb)	1000	Sin dato
Longitud del fragmento más grande (pb)	441043	374700

Para completar el ensamblado se plantearon diversos procedimientos que permitiesen elucidar el orden y la orientación de los fragmentos, de forma que se organizaran en secuencias de mayor tamaño. Las evidencias en las que se basaron las hipótesis de unión de los fragmentos fueron las siguientes: (1) la presencia de una secuencia proteica truncada entre los extremos de los fragmentos, (2) el orden de los genes en organismos cercanos filogenéticamente a TFA, como *Sphingopyxis alaskensis* RB2256 y (3) el solapamiento de la secuencia de ADN entre los extremos de algunos de los fragmentos. Este solapamiento, en algunos casos, ocurría entre regiones repetidas presentes en los extremos de más de dos fragmentos, lo cual no podía resolverse con los ensambladores automáticos.

## Resultados

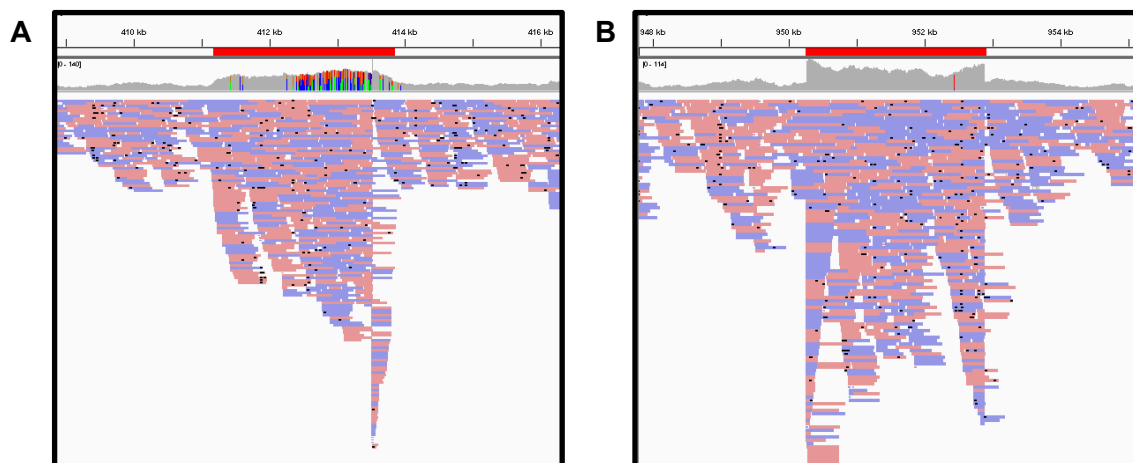
Una vez establecido el posible orden entre los fragmentos para los cuales se encontró alguna evidencia de unión, se diseñaron oligonucleótidos en sus extremos, orientados hacia fuera, para amplificar la zona por PCR y secuenciarla.

Se consiguieron 30 uniones gracias a las hipótesis planteadas anteriormente y a su posterior confirmación. Con la hipótesis 1 se redujo el número de fragmentos de 42 a 37; con la 2, de 37 a 32 y con la 3, de 32 a 12. Sin embargo, para la unión de esos 12 fragmentos, no se pudieron plantear hipótesis de unión o estas no pudieron ser confirmadas, por lo que fue necesario utilizar las dos estrategias siguientes: (1) Identificación de cósmidos en una genoteca de TFA, mediante *Southern Blot* o PCR, que contuviesen el extremo de uno de los fragmentos (ver apartado 2.2 de materiales y métodos) y (2) realización de PCRs combinando oligonucleótidos que hibridaban en los extremos de los fragmentos hasta encontrar las parejas que amplificasen la unión entre dos fragmentos (PCRs aleatorias). Se redujo el número de fragmentos de 12 a 9 gracias a los experimentos de *Southern Blot* y de 9 a 5 por la realización de PCRs aleatorias. La identificación de cósmidos por PCR de la secuencia del extremo de los fragmentos permitió confirmar una nueva unión, disminuyendo el número de fragmentos a 4.

Aunque con la metodología seguida hasta el momento se cubrían todas las opciones de unión, no fue posible cerrar completamente el genoma. Se obtuvo un genoma parcial constituido por 4 grandes fragmentos (ctg149 de 4 kb, ctg220-217 de 1,8 Mb, ctg223ic-219ic de 1,7 Mb y ctg229-248 de 1 Mb) sobre los que se alinearon las pequeñas lecturas obtenidas de la secuenciación inicial, utilizando el software BWA (*Burrows-Wheeler Aligner*; Li y Durbin, 2009). Para ello, se generó un fichero FASTA que contenía un genoma artificial con los 4 fragmentos unidos de forma consecutiva, en el orden en el que se nombraron anteriormente. Este fichero y el fichero FASTA donde se almacenan las lecturas de la secuenciación se utilizaron como ficheros de entrada para la herramienta BWA, con la que se llevó a cabo el alineamiento de las pequeñas lecturas en el genoma artificial. El alineamiento se almacenó finalmente en formato SAM (Li y col., 2009) y el resultado del alineamiento se visualizó a través de IGV (*Integrative Genome Viewer*; Thorvaldsdóttir y col., 2013) y se analizó la cobertura de la secuenciación a lo largo del genoma artificial.

Esta nueva aproximación permitió identificar dos regiones del genoma con una gran acumulación de lecturas dentro del fragmento ctg220-217. Una de ellas contenía las secuencias de tres transposasas consecutivas (Tp123) y una resolvasa (R) (Figura19A) y la otra, localizada unas 536 kb aguas abajo de la anterior, los genes *istA* e *istB* que constituyen, respectivamente, las subunidades catalítica y de unión a ATP de una integrasa (I) (Figura19B).

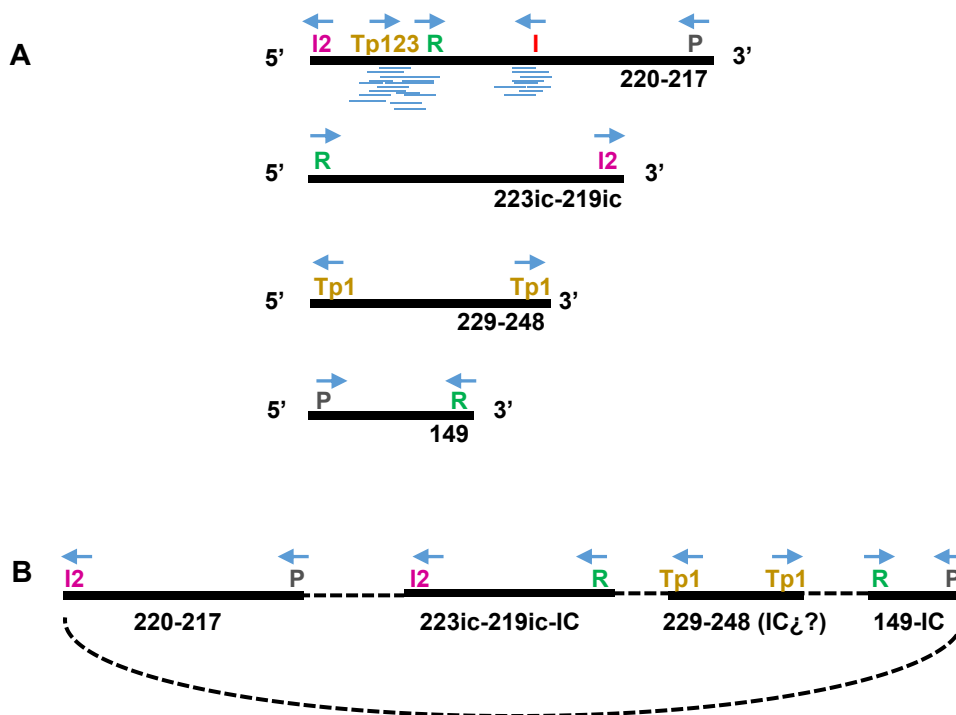
Además, parte de esas regiones se repetían en algunos de los extremos de los 4 fragmentos que aún quedaban por ensamblar. Por tanto, los huecos no representados entre los fragmentos podrían corresponder a repeticiones de las zonas del genoma artificial donde se detectaron acumulaciones de lecturas.



**Figura 19. Acumulación de lecturas en el genoma artificial de TFA, las cuales se localizan en el fragmento ctg220-217.** En rojo se delimitan las regiones con lecturas acumuladas correspondientes a genes codificantes para 3 transposasas consecutivas y una resolvasa, Tp123 y R (A) y codificantes para las subunidades IstA e IstB de una integrasa, I (B). Las lecturas coloreadas en rosa alinean en sentido *forward* y las azules en sentido *reverse*. En gris se muestra la cobertura del genoma por parte de las lecturas y las marcas de colores en dicha franja gris indican que se podrían encontrar varias bases en una determinada posición.

Parte de la secuencia de Tp1 también se encontró en ambos extremos de otro de los fragmentos, ctg229-248, en direcciones opuestas. Un trozo de la secuencia de la resolvasa se repetía en el extremo 5' del fragmento ctg223ic-219ic y en el extremo 3' del ctg149, en direcciones opuestas. La región codificante para una nueva integrasa (I2) aparecía en el extremo 5' del fragmento ctg220-217, así como en el 3' del ctg223ic-219ic, en direcciones opuestas (Figura 20A). Por último, en los extremos 3' de ctg220-217 y 5' de ctg149, se localizó parte de la región codificante para una primasa (P).

## Resultados



**Figura 20. Representación esquemática de las acumulaciones de lecturas en el fragmento ctg220-217 y de las repeticiones de secuencias en los extremos de los 4 fragmentos finales del genoma de TFA.** (A) Se muestran las proteínas codificadas en los extremos de los cuatro fragmentos así como la acumulación de lecturas en ctg220-217. (B) Modelo planteado para finalizar el ensamblaje del genoma. En el esquema, IC denota que se trata de la inversa complementaria del fragmento grande e ic del fragmento inicial y/o final que se indica. Las distintas letras representan: R, resolvasa; I, integrasa; I2, integrasa 2; Tp123, transposasas 1, 2 y 3; Tp1, transposasa 1 y P, primasa.

Para unir los 4 fragmentos finales y terminar el ensamblaje del genoma, se planteó el modelo de unión que se representa en la Figura 20B, basado en las observaciones descritas anteriormente. El extremo 3' de la reversa complementaria del fragmento ctg223ic-219ic podría ir unido tanto con el fragmento ctg229-248 como con su reversa complementaria y este, a su vez, con el extremo 5' de la reversa complementaria del ctg149, de manera que quedarán próximas la transposasa Tp1 y la resolvasa R, como ocurre en el fragmento ctg220-217. Considerando la acumulación de lecturas en la integrasa I dentro el genoma artificial, se planteó que en los huecos entre los fragmentos anteriores se encontrase dicha integrasa, quedando localizada entre las transposasas y la resolvasa.

Al analizar la conservación de la primasa P en otras bacterias, se detectó que se encontraba muy próxima a la integrasa I2, como ocurre en el plásmido pNL1 de *Novosphingobium aromaticivorans* DSM 12444. La integrasa I2 de TFA alinea con una similitud del 93 %, una identidad del 88 % y un e-value de 0 con la integrasa I2 de *N.*

*aromaticivorans* DSM 12444 (Identificador Uniprot: A4XE10), mientras que el fragmento de la primasa P que se encuentra en los extremos 3' de ctg220-217 y el 5' de la reversa complementaria del fragmento ctg149 alinea con el final de una proteína no caracterizada de *N. aromaticivorans* DSM 12444, la cual, según el alineamiento realizado, podría tratarse de una primasa (Identificador Uniprot: A4XE13). Los genes que codifican para estas proteínas en *N. aromaticivorans* DSM 12444 están separados por solo dos secuencias codificantes. Se planteó que en TFA existiese una organización parecida, de forma que el extremo 5' del fragmento ctg220-217 iría unido con el extremo 3' de la reversa complementaria de ctg149 y el otro extremo del fragmento ctg220-217 con el 5' de la reversa complementaria de ctg223ic-219.

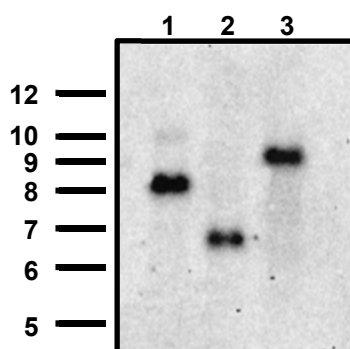
Para terminar de ensamblar el genoma, siguiendo el modelo de la Figura 20B, era necesario conocer la secuencia de los huecos entre los fragmentos así como determinar el orden correcto del fragmento ctg229-248. Para el primer objetivo, se localizaron las lecturas del final de los fragmentos en el archivo SAM generado por BWA y se fueron ensamblando por sus extremos 3' con lecturas que, según la información proporcionada por la empresa LifeSequencing S.L., no habían sido asignadas a ninguno de los 42 fragmentos iniciales y debían corresponder a las secuencias de los huecos entre ellos. A su vez, las lecturas no asignadas se ensamblaron entre sí hasta rellenar los huecos y alcanzar el extremo de un nuevo fragmento. De esta forma, se validó la predicción del modelo de la Figura 20B y se obtuvieron las secuencias de los huecos entre los 4 fragmentos.

Por último, se elucidó la orientación del fragmento 229-248 a través de un experimento de *Southern Blot* (Figura 21). Se hibridó una sonda del extremo 5' del fragmento 229-248 en una membrana de nailon que contenía separados los fragmentos de restricción del ADN genómico de TFA con las enzimas HindIII, EcoRI y BamHI. Se determinó *in silico* el tamaño esperado de las bandas que hibridarían con la sonda, tanto en la orientación fijada en el modelo de la Figura 20B (ctg229-248), como en la reversa complementaria (ctg229-248-IC). Para las restricciones con EcoRI y BamHI se esperaba obtener, respectivamente, un fragmento de 6,5 kb y de 9,6 kb, independientemente de la orientación de ctg229-248, y en la restricción con HindIII un fragmento de 16,7 kb para la orientación actual del fragmento ctg229-248 y de 8,5 kb para su reversa complementaria.

El tamaño de la banda de la restricción con HindIII en la que hibridase la sonda permitiría conocer la orientación. A pesar de que las bandas de hibridación de las otras restricciones no mostrarían diferencia entre las orientaciones, se realizaron para confirmar que la secuencia de los gaps predicha *in silico* era la correcta. El resultado indicó que la orientación de este

## Resultados

fragmento era su reversa complementaria, como puede verse en la Figura 21 (ctg229-248-IC).



**Figura 21. Resultado del Southern Blot empleando la sonda del extremo 5' del fragmento ctg229-248 sobre digestiones del ADN cromosómico de TFA.** En el carril 1 se representa el resultado de la hibridación en la restricción con HindIII, en el carril 2 en la restricción con EcoRI y en el carril 3 en la restricción con BamHI. A la izquierda se muestra el patrón de peso molecular (kb).

Aunque las regiones repetidas han dificultado el ensamblaje, constituyendo el cuello de botella en el proceso de finalización del genoma y obligando a realizar estrategias alternativas que permitiesen resolverlas, los 42 fragmentos iniciales proporcionados por la empresa LifeSequencing S.L. pudieron ensamblarse en una única secuencia circular de 4679853 pb que conforma el genoma completo de TFA.

### 1.2. Características generales del genoma de *S. granuli* TFA

El genoma de TFA completamente ensamblado constituye una única molécula de ADN circular de aproximadamente 4,7 Mb (4679853 pb) y presenta un contenido en G+C del 66,2 %. Este alto porcentaje en G+C concuerda con el establecido para el género *Sphingopyxis* (62,3-69,2 %) por Glaeser y Kämpfer (2014). La secuencia genómica obtenida fue anotada a través de diversas herramientas bioinformáticas. Inicialmente, se delimitaron las pautas abiertas de lectura de los genes codificantes para proteínas con el servidor Prodigal v1.20 (Hyatt y col., 2010).

A continuación, a cada una de las pautas abiertas de lectura definidas se les asignó una anotación funcional con la herramienta Sma3s, tal y como se describe en materiales y métodos. Sma3s es un programa de anotación automática escrito en lenguaje de programación Perl que se basa en la similitud de las proteínas que se quieren anotar con las almacenadas en la base de datos UniProt (Swiss-Prot/TrEMBL). En el caso de la anotación del genoma de TFA, la base de datos de UniProt se redujo de forma que sólo se tuvieron en

cuenta las proteínas anotadas en bacterias y no la base de datos completa. Esta herramienta ha mostrado una elevada precisión en la asignación de funciones a proteínas bacterianas (Muñoz-Mérida y col., 2014). Además, para la anotación del genoma de TFA se priorizaron las anotaciones provenientes de la base de datos Swiss-Prot (revisadas) frente a las de TrEMBL (no revisadas) con el fin de asignar información de mayor calidad. Para finalizar con la anotación de las proteínas se llevó a cabo la asignación a las categorías COG mediante la utilización de la herramienta WebMGA (Wu y col., 2011). Posteriormente, se anotaron los ARNs funcionales, constituidos por los ARNs transferentes (ARNt), ARNs ribosómicos (ARNr) y ARNs reguladores no codificantes. Para ello, se utilizaron herramientas bioinformáticas especializadas en la anotación de cada uno de ellos (ver apartado 3.1 de materiales y métodos).

La secuencia y la anotación, revisadas manualmente, se subieron al *National Centre for Biotechnology Information* (NCBI) bajo el BioProject PRJNA283604 y el número de acceso a GenBank CP012199. Como resultado de la anotación del genoma se predijeron 4250 genes, tanto codificantes para proteínas como para ARNs funcionales, lo cual representa un 89,47 % de región codificante con respecto al total de la secuencia genómica. Las características generales del genoma de TFA y un resumen de la anotación funcional se muestran en la Tabla 9.

**Tabla 9. Características de la secuencia y anotación del genoma de *Sphingopyxis granulii* TFA.**

Característica	Valor
Tamaño del genoma (pb)	4679853
Contenido en G+C (%)	66,2
Secuencia codificante (pb)	4187393
Total de genes predichos	4250
Genes codificantes para proteínas	4190
Proteínas con descripción asignada	3502
Proteínas no caracterizadas	688
Número de ARNs no codificantes	60
ARNt	46
ARNr	3
Otros ARNs no codificantes	11

## Resultados

Un 16,4 % del total de proteínas definidas en el genoma de TFA se anotaron como proteínas de función desconocida, lo cual fue debido a la falta de similitud entre ellas y las proteínas contenidas en las bases de datos o a la similitud con proteínas no caracterizadas de otras bacterias.

En cuanto a los ARN funcionales, los 3 ARN ribosómicos anotados corresponden a las subunidades 16S, 23S y 5S de la fracción ribonucleica que constituyen los ribosomas en bacterias. El hecho de que en TFA sólo se haya detectado un operón ribosómico concuerda con lo establecido para bacterias de crecimiento lento y con un estilo de vida oligotrofo (Lauro y col., 2009). Por otra parte, los 46 ARN transferentes localizados son suficientes para la lectura de todos los codones de TFA (ver apartado siguiente apartado), lo cual constituye un control de calidad de la anotación de los mismos.

### 1.2.1. Análisis del uso de codones

Con los 46 ARNt detectados en TFA y atendiendo al código genético, no se cubría la decodificación de todos los codones posibles para la síntesis de proteínas. Sin embargo, la regla del tambaleo codón-anticodón explicaría la lectura de los codones sin anticodón asociado (Tabla 10 y Tabla 11). Esta regla establece que existe un reconocimiento imperfecto entre la tercera base del codón y la primera base del anticodón pero no de forma azarosa, sino de acuerdo con unas normas y con la intervención de nucleósidos modificados en la primera posición de algunos anticodones (Tabla 10) (Murphy y Ramakrishnan, 2004).

Tabla 10. Regla del tambaleo codón-anticodón.

Base 5' en el anticodón del ARNt	Base 3' en el codón del ARNm
G	C, U
C	G
A	U
U	A, G
I (inosina)	A, C, U
L (lisidina)	A

Para la lectura del codón CGA que codifica para la arginina (Arg) no se encontró ningún ARNt y, según el tambaleo clásico permitido (considerando A, C, U, G), ninguno de los otros ARNt<sup>Arg</sup> presentes en TFA podrían codificarlo. Se ha descrito que la adenosina de la primera base del anticodón ACG (ARNt<sup>Arg</sup><sub>ACG</sub>) se modifica a inosina (I) mediante una reacción de desaminación llevada a cabo por la enzima adenosina desaminasa (TadA) (Wolf y col., 2002),



codificada por el gen SGRAN\_0704 en TFA. La inosina es estructuralmente parecida a la guanina pero le falta del grupo 2-amino. Como consecuencia, forma pares de bases con U y C como la guanina y, adicionalmente, con A, lo cual no puede hacer la guanina debido a la presencia del grupo 2-amino. Por tanto el codón CGA podría leerse en TFA por ARNt<sup>Arg</sup><sub>ICG</sub>, al igual que el codón CGC (Tabla 11).

Tampoco se encontró un ARNt para el codón AUA que codifica para la isoleucina. Sin embargo, la proteína ARNt(Ile)-2-lisil-citidina sintasa (TlIS) codificada por el gen SGRAN\_2051 en TFA, cataliza la ligación del aminoácido lisina a la citidina de la primera base del anticodón CAU del ARNt<sup>Met</sup>, de forma que se convierte en lisidina (L). Este nucleósido forma pares de bases con A en lugar de con G, permitiendo discriminar entre el codón AUG (metionina) y AUA (isoleucina). Además, el cambio entre C y L hace que el mismo ARNt cargue metionina o isoleucina, respectivamente (Voorhees y col., 2013).

**Tabla 11. Uso de codones en *Sphingopyxis granuli* TFA.** En rojo se indican las bases implicadas en el tambaleo entre codón-anticodón y entre paréntesis los nucleósidos modificados. En el caso de la serina se distinguen distintos tambaleos en color rojo y azul.

Aminoácido	Codón (5' → 3')	Cantidad	Anticodón (3' → 5')	Nº ARNt	Modo de lectura	% uso codón	% uso codón por aa
Ninguno (STOP)	UAA	496	--	--	--	0,04	--
Ninguno (STOP)	UAG	783	--	--	--	0,06	--
Ninguno (STOP)	UGA	2899	--	--	--	0,21	--
Alanina	GCU	7482	--	0	Tambaleo	0,54	3,87
	GCA	11455	CGU	1	ARNt	0,82	5,93
	GCC	72364	CGG	1	ARNt	5,21	37,47
	GCG	101827	CGC	1	ARNt	7,33	52,73
Arginina	AGA	998	UCU	1	ARNt	0,07	0,94
	AGG	3104	UCC	1	ARNt	0,22	2,92
	CGA	4180	GC(I)	0	ARNt <sup>Arg</sup> <sub>ICG</sub>	0,30	3,93
	CGU	7478	GCA	1	ARNt	0,54	7,02
	CGG	25089	GCC	2	ARNt	1,81	23,57
	CGC	65608	GC(I)	0	ARNt <sup>Arg</sup> <sub>ICG</sub>	4,72	61,63
Asparagina	AAU	11245	--	0	Tambaleo	0,81	33,79
	AAC	22037	TTG	1	ARNt	1,59	66,21
Aspartato	GAU	25793	--	0	Tambaleo	1,86	30,04
	GAC	60079	CUG	1	ARNt	4,32	69,96
Cisteína	UGU	1271	--	0	Tambaleo	0,09	11,93

## Resultados

Aminoácido	Codón (5'→ 3')	Cantidad	Anticodón (3'→ 5')	Nº ARNt	Modo de lectura	% uso codón	% uso codón por aa
	UGC	9387	ACG	1	ARNt	0,68	88,07
Glutamato	GAA	31766	CUU	1	ARNt	2,29	43,31
	GAG	41573	CUC	1	ARNt	2,99	56,69
Glutamina	CAA	5782	GUU	1	ARNt	0,42	14,07
	CAG	35319	GUC	1	ARNt	2,54	85,93
Glicina	GGA	9769	CCU	1	ARNt	0,70	7,83
	GGU	10020	--	0	Tambaleo	0,72	8,03
	GGG	23344	CCC	1	ARNt	1,68	18,71
	GGC	81659	CCG	1	ARNt	5,88	65,44
Histidina	CAU	13058	--	0	Tambaleo	0,94	47,72
	CAC	14307	GUU	1	ARNt	1,03	52,28
Isoleucina	AUA	2368	UA(L)	0	ARNt <sup>lle</sup> <sub>LAU</sub>	0,17	3,44
	AUU	6455	--	0	Tambaleo	0,46	9,37
	AUC	60074	UAG	1	ARNt	4,32	87,19
Leucina	UUA	552	AAU	1	ARNt	0,04	0,41
	CUA	1777	GAU	1	ARNt	0,13	1,31
	UUG	8080	AAC	1	ARNt	0,58	5,96
	CUU	10535	--	0	Tambaleo	0,76	7,76
	CUC	47723	GAG	1	ARNt	3,43	35,17
	CUG	67012	GAC	1	ARNt	4,82	49,39
Lisina	AAA	6634	AAA	1	ARNt	0,48	17,26
	AAG	31808	AAG	1	ARNt	2,29	82,74
Metionina	AUG	32162	UAC	3	ARNt	2,31	100,00
Fenilalanina	UUU	7032	--	0	Tambaleo	0,51	14,48
	UUC	41524	AAG	1	ARNt	2,99	85,52
Prolina	CCA	2864	GGU	1	ARNt	0,21	3,83
	CCU	3260	--	0	Tambaleo	0,23	4,36
	CCC	26790	GGG	1	ARNt	1,93	35,84
	CCG	41829	GGC	1	ARNt	3,01	55,96
Serina	UCU	1564	--	0	Tambaleo	0,11	2,22
	UCA	1726	AGU	1	ARNt	0,12	2,45
	AGU	2429	--	0	Tambaleo	0,17	3,45
	UCC	9162	AGG	1	ARNt	0,66	13,03
	AGC	25849	UCG	1	ARNt	1,86	36,75
	UCG	29603	AGC	1	ARNt	2,13	42,09
Treonina	ACU	2240	--	0	Tambaleo	0,16	3,12
	ACA	2964	UGU	1	ARNt	0,21	4,13
	ACG	29876	UGC	1	ARNt	2,15	41,67
	ACC	36622	UGG	1	ARNt	2,64	51,08
Triptófano	UGG	20083	ACC	1	ARNt	1,45	100,00

Aminoácido	Codón (5' → 3')	Cantidad	Anticodón (3' → 5')	Nº ARNt	Modo de lectura	% uso codón	% uso codón por aa
Tirosina	UAC	9151	AUG	1	ARNt	0,66	29,55
	UAU	21819	--	0	Tambaleo	1,57	70,45
Valina	GUA	2311	CAU	1	ARNt	0,17	2,37
	GUU	6582	--	0	Tambaleo	0,47	6,74
	GUG	32541	CAC	1	ARNt	2,34	33,33
	GUC	56186	CAG	1	ARNt	4,04	57,56

El mecanismo de tambaleo se requiere para la lectura de al menos un codón de cada aminoácido, excepto en el caso del glutamato, la glutamina, la lisina, la metionina y el triptófano, los cuales disponen de todos los ARNt para su lectura. Además, los codones más abundantes para arginina (CGC) y tirosina (UAU) se leen gracias a este mecanismo. Por otra parte, no se han detectado en el genoma de TFA ni genes codificantes de ARNt para la selenocisteína ni posibles selenoproteínas.

### 1.2.2. Identificación del origen de replicación

El origen de replicación de TFA, OriC, se detectó a través del programa Ori-Finder (Gao y Zhang, 2008). Como entrada a la herramienta se utilizó el fichero FASTA que contenía la secuencia completa del genoma de TFA y un fichero con la anotación en formato PTT. La localización de las cajas DnaA se realizó teniendo en cuenta el consenso de una caja perfecta de *E. coli* (TTATCCACA) permitiendo solo un error.

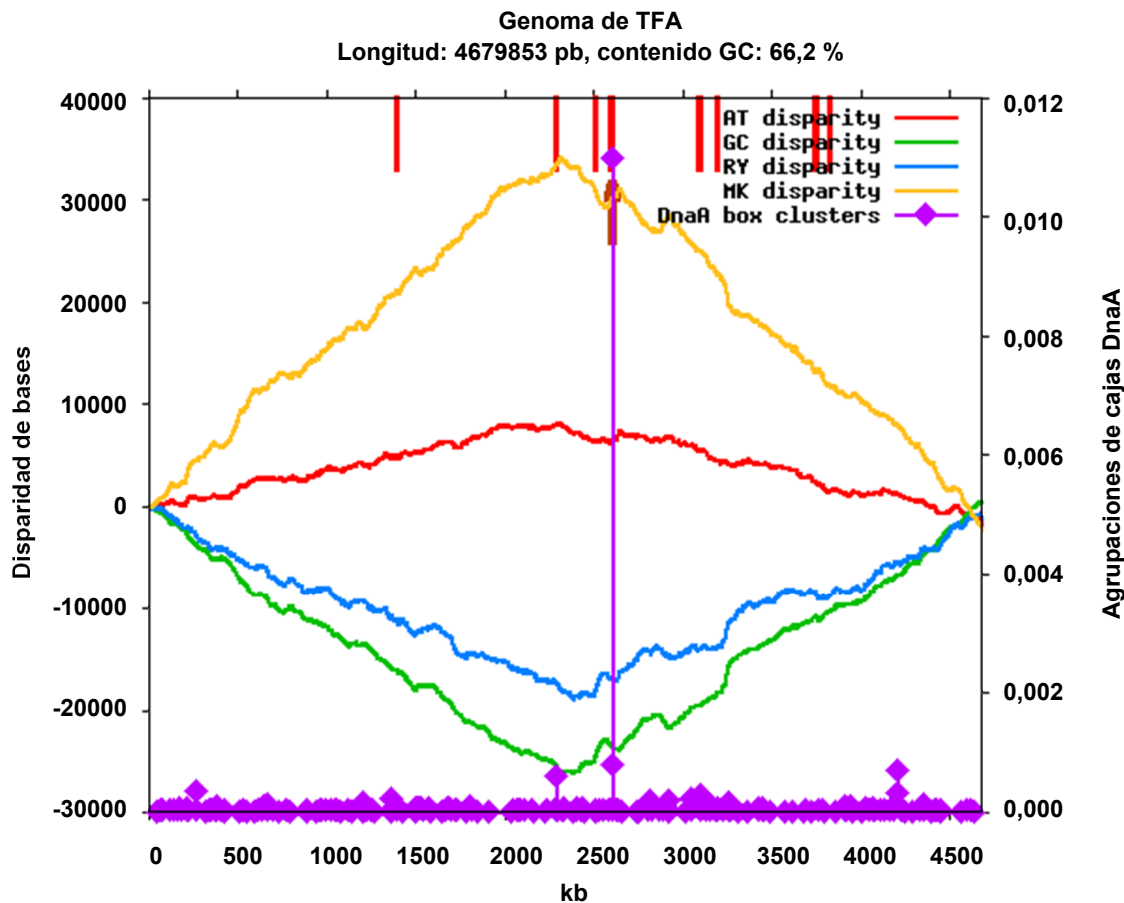
La secuencia del OriC predicha por Ori-Finder, con una longitud de 335 nucleótidos, se muestra en la Figura 22A y el análisis de los elementos en los que se basa la predicción se representa en la Figura 22B.

## Resultados

A

```
GCGGCCCTCTTTCTTCTATCAAAACAAGTATAAATGGATTGTGGTGGATTGTTGGCGGGCGGATAAGTGCGCTT
TAGGCGCAGGCGCCCTTTTGCCAACATCTTGACTCCCCGGCCGCCCGTCCGCCGCGGAGTCGGTGGCGGTCG
TCCCCGTTCTTACAGCCTGTGGACAAGATTTCGTCCCCTGTGGATAAACGGCGGGACGGAATCGGCGCGGCCGGG
GATGGAATCGCTGTCCCGCTTCGCCCCCGGCTTGTGCCGGACTTTGTCCACAGCCGGCCCGCGATGCTTCGCCGC
TTGCCTTTTCGGGCGGCGCGGGGACAAGAAGGGCG
```

B



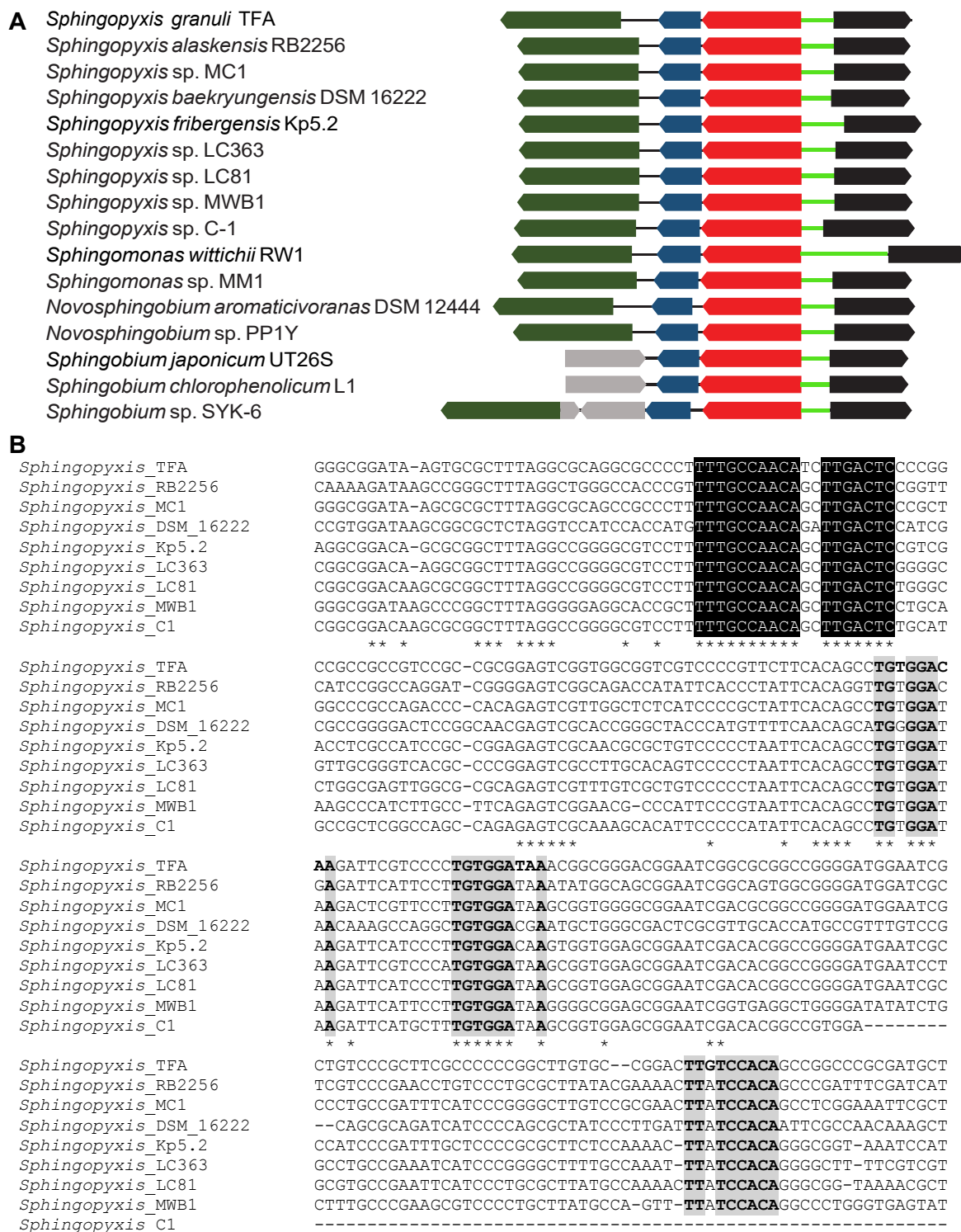
**Figura 22. Resultado de la predicción de Ori-Flnder.** (A) Secuencia del OriC predicho en el genoma de TFA. Se marcan en negrita y subrayadas las cajas DnaA identificadas y en rojo las bases que no coinciden con la secuencia de la caja DnaA de *E. coli*. (B) Z-curvas (AT, GC, RY y MK) de la secuencia rotada del genoma que comienza y termina en el máximo de la curva de disparidad GC. Las líneas rojas verticales señalan la localización de genes indicadores y la flecha vertical marrón la localización del origen de replicación. Los picos morados con los rombos indican las agrupaciones de cajas DnaA.

Se han localizado 3 cajas DnaA en la región definida como OriC, las dos primeras en el sentido reverso a la secuencia y la tercera en el sentido de la secuencia (Figura 22A). La localización del origen de replicación suele estar próxima a las regiones donde se producen cambios en la distribución de las bases (Zhang y Zhang, 2014), como ocurre en TFA, cuyo OriC se encuentra próximo al cambio de pendiente de las Z-curvas (Figura 22B). Además, en TFA, el OriC se localiza cerca del gen *hemE* (SGRAN\_0274), como en la mayoría de las bacterias de la familia *Sphingomonadaceae* incluidas en la base de datos DoriC (Gao y col., 2013), excepto en *Sphingopyxis alaskensis* RB2256. En los siguientes análisis, para esta bacteria se tomó la región intergénica homóloga a la que contiene el OriC en TFA y en el resto de *Sphingomonadaceae*.

La comparación de la región genómica del OriC mostró una conservación del contexto génico en la familia *Sphingomonadaceae* (Figura 23A). Además, el alineamiento a través de Clustal Omega de la región intergénica que contiene el origen de replicación en las distintas *Sphingopyxis* (Figura 23B) puso de manifiesto una alta conservación de los sitios característicos de los orígenes de replicación, como son las cajas DnaA y el sitio de desenrollamiento del ADN (*DNA unwinding element*, DUE) (Ozaki y Katayama, 2012).

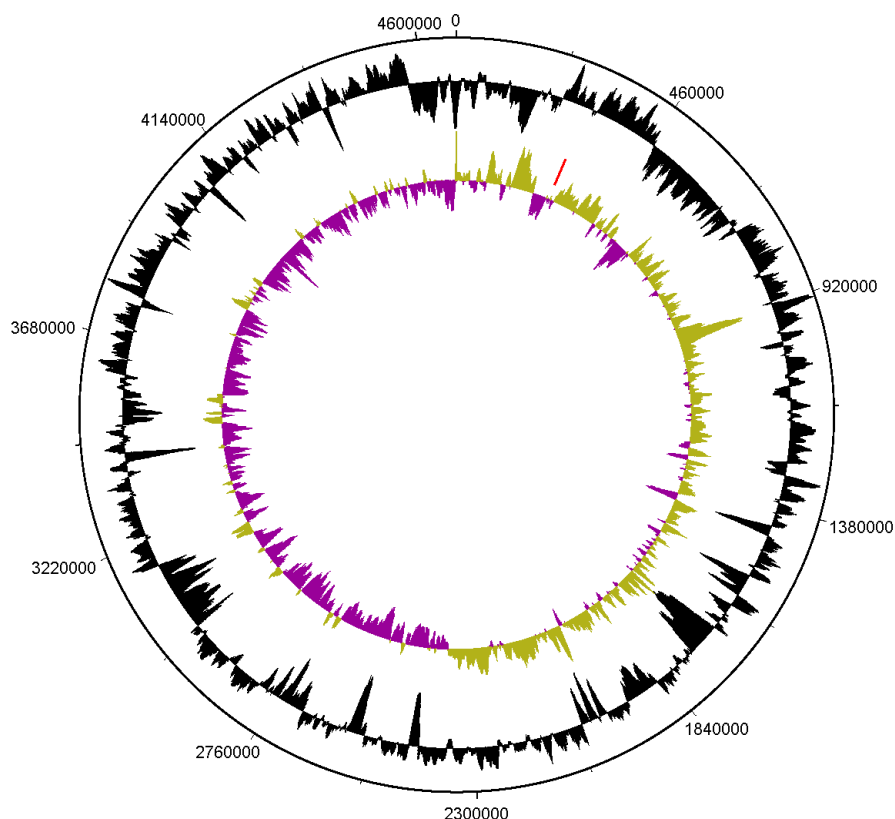
Tanto la conservación del contexto génico como de los sitios característicos del OriC sugieren que el origen de replicación de *S. alaskensis* RB2256 no está localizado correctamente en la base de datos DoriC, y se situaría en la localización representada en la Figura 23A, al igual que en el resto de bacterias estudiadas.

## Resultados



**Figura 23. Análisis de la región genómica que contiene el OriC.** (A) Organización génica alrededor del OriC en representantes de la familia *Sphingomonadaceae*. El OriC (raya verde limón) se encuentra entre los genes SGRAN\_0275 (en negro) que codifica para una proteína reguladora y SGRAN\_0274 (*hemE*, en rojo) que codifica para una uroporfirinógeno descarboxilasa. También se conservan un gen codificante para una proteína de membrana (SGRAN\_0274, en azul) y el factor Rho (SGRAN\_0272, en verde). En gris se representan genes sin sintenia. (B) Alineamiento con Clustal Omega de la zona donde se conservan los elementos típicos de los OriC de las distintas *Sphingopyxis*. Las posibles cajas DnaA, predichas por OriFinder, se muestran en gris y el sitio DUE en negro. Los asteriscos muestran la conservación de los nucleótidos.

Otro parámetro utilizado tradicionalmente para la localización del origen de replicación es la desviación GC a lo largo de la secuencia genómica (Xia, 2012). Se usó el programa DNAPlotter v1.11 (Carver y col., 2009) para representar la desviación GC en el genoma de TFA. La desviación GC separa al genoma de TFA en dos regiones bien diferenciadas, una con desviación GC por encima de la media y otra por debajo de la media (Figura 24).



**Figura 24. Representación del contenido en GC (en negro) y la desviación GC, (en verde y morado), en el genoma de TFA.** En ambos casos, hacia fuera del círculo se representa los valores por encima de la media y hacia dentro por debajo de la media. Se indica en rojo la localización del origen de replicación OriC predicho por OriFinder.

En general, en los genomas con mecanismo de replicación bidireccional, en la cadena líder hay un enriquecimiento de G frente a C con respecto a la cadena retardada (Xia, 2012), por lo que, en el genoma de TFA, la región verde de la Figura 24 correspondería a la cadena líder y la morada a la retardada. Además, las regiones de inicio y fin de la replicación coinciden con los puntos de inflexión en la desviación GC (Worning, 2006). Atendiendo a lo anterior, el origen de replicación de TFA se localizaría en torno a la base 0 representada en la Figura 24. Sin embargo, se observa que el OriC detectado por Ori-Finder para TFA está desplazado de la inflexión en la desviación GC.

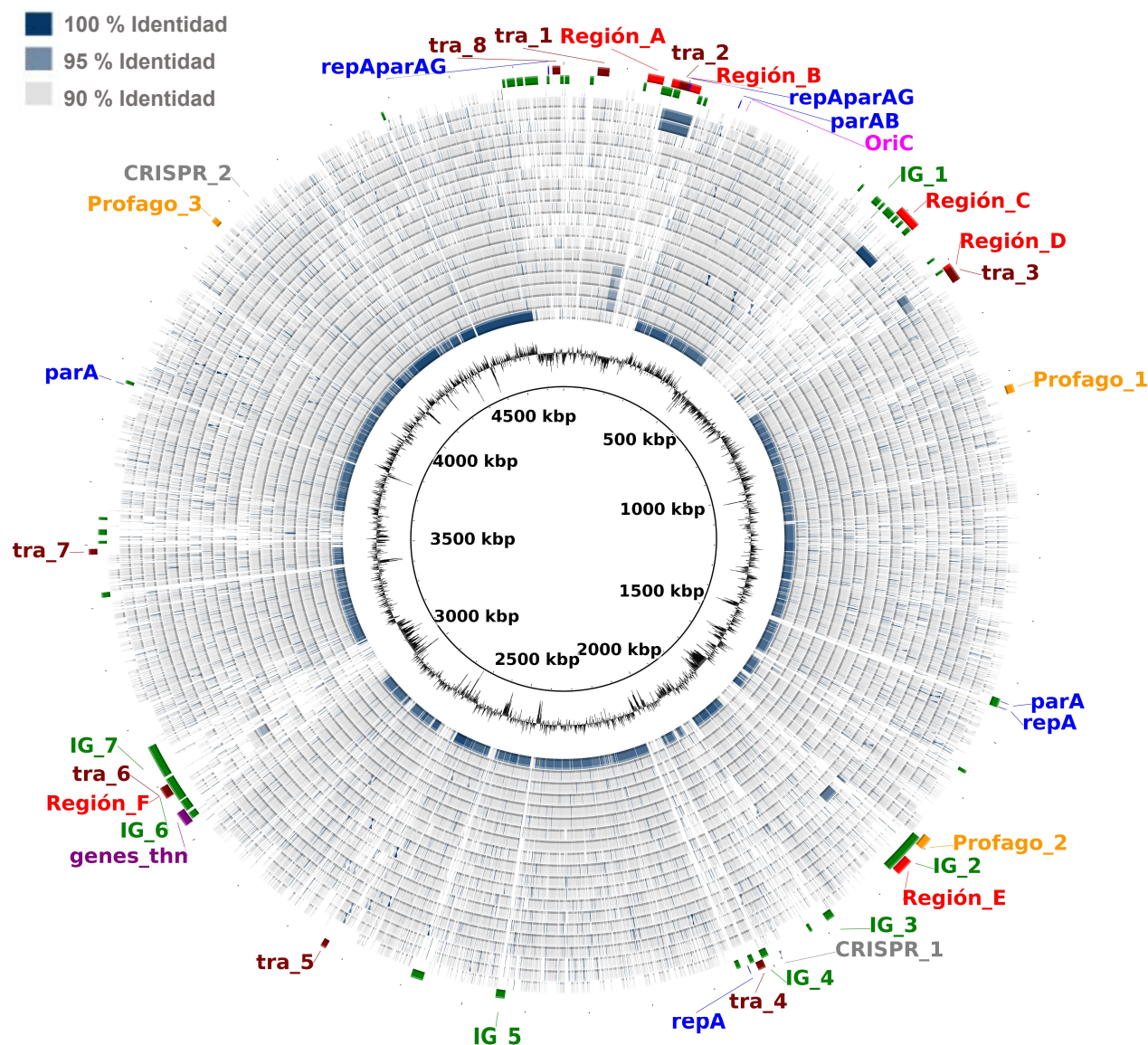
### 1.3. Organización del genoma de *S. granuli* TFA: evidencias de transferencia horizontal, islas genómicas y profagos

#### 1.3.1. Identificación de regiones del genoma de *S. granuli* TFA homólogas a otros genomas.

La secuencia del genoma de TFA se comparó mediante BLASTN con genomas de bacterias pertenecientes a la clase  $\alpha$ -proteobacteria (Figura 25) a través de la herramienta BRIG (*BLAST Ring Image Generator*) (Alikhan y col., 2011). Gracias a dicha comparación, se definieron seis regiones con una gran similitud de secuencia entre TFA y algunos de los genomas alineados (anillo 23 de la Figura 25). La región A (coordenadas 134249-159746) presenta una identidad del 95 % con regiones en los cromosomas de *Sphingopyxis* sp. MC1, *S. alaskensis* RB2256, *S. fribergensis* Kp5.2 y *Sphingopyxis* sp. LC363 y contiene principalmente genes implicados en la resistencia a metales pesados. La región B (coordenadas 173796-220082) es un 97 % idéntica a una parte del genoma de *Oligotropha carboxidovorans* OM4/OM5 y contiene genes relacionados con el transporte de potasio, la conjugación, replicación y partición de plásmidos, el metabolismo del ADN y otras proteínas no caracterizadas. Las regiones C y D muestran un 99 y 96 % de identidad, respectivamente, con regiones del cromosoma de *Sphingobium chlorophenolicum* L1. En la región C (coordenadas 590566-628642) están presentes proteínas implicadas en detección de oxígeno y respuesta a estrés mientras que en la D (coordenadas 707273-722772) se encuentran genes que codifican para proteínas de transferencia conjugativa de plásmidos. La región E (coordenadas 1728811-1756443) es 97 % idéntica a una región del genoma de *Erythrobacter litoralis* HTCC2594 que presenta genes codificantes para dioxigenasas, monooxigenasas, hidrolasas y sistemas de transporte implicados en resistencia a drogas. Por último, la región F (coordenadas 3076722-3095970) es muy similar, con un 94 % de identidad, a una zona del cromosoma de *Novosphingobium* sp. estirpe PP1Y que engloba genes relacionados con conjugación de plásmidos.

En cuanto a la comparación con *Sphingopyxis granuli* ku-sg, el porcentaje de identidad a lo largo de todo el genoma presenta valores muy próximos al 100 %, siendo la bacteria cuyo genoma es el más similar al de TFA de todos los incluidos en la comparación.





**Figura 25. Representación circular del genoma de TFA y comparación con los genomas de otras  $\alpha$ -proteobacteria.** La localización del OriC se marca en rosa. La comparación entre el genoma de TFA y los genomas de varias  $\alpha$ -proteobacterias se realizó a través de BLASTN con la herramienta BRIG. Los 23 anillos representan, desde dentro hacia fuera: (1) Contenido en GC, (2-21) Representación de la similitud entre el genoma de TFA y los genomas de *Sphingopyxis granuli* ku-sg, *Sphingopyxis* sp. MC1, *Sphingopyxis alaskensis* RB2256, *Sphingopyxis fribergensis* Kp5.2, *Sphingopyxis* sp. LC363, *Sphingopyxis* sp. LC81, *Sphingopyxis* sp. MWB1, *Sphingopyxis* sp. C-1, *Sphingopyxis baekryungensis* DSM 1622, *Sphingomonas wittichii* RW1, *Novosphingobium aromaticivorans* DSM 12444, *Novosphingobium* sp. PP1Y, *Erythrobacter litoralis*, *Blastomonas* sp. AAP53, *Sphingobium* sp. SYK-6, *Sphingobium japonicum* UT26, *Sphingobium chlorophenolicum* L-1, *Oligotropha carboxidovorans* OM4, *Oligotropha carboxidovorans* OM5 y *Zymomonas mobilis*; (22) Islas Genómicas predichas por IslandViewer 3 (regiones verdes) y, dentro de ellas, se numeran como IG\_1 a 7 las predichas por IslandPath-DIMOB; (23) Genes *tra/trb* (*tra*\_1 al 8, en marrón), regiones con alta similitud entre TFA y otras  $\alpha$ -proteobacteria, excluyendo a *Sphingopyxis granuli* ku-sg, (Región\_A a la F, en rojo), genes de replicación y partición de plásmidos (azul), profagos predichos (Profago\_1 al 3, en amarillo), secuencias CRISPR (CRISPR 1 y 2, en gris) y los genes *thn* (en morado).

## Resultados

### 1.3.2. Identificación de islas genómicas

Las islas genómicas (IGs) son grandes regiones del cromosoma que presentan una composición de bases diferente con respecto al resto del genoma. Estos fragmentos son adquiridos por transferencia horizontal y suelen contener genes no esenciales que aportan una gran diversidad de funciones, como la utilización de nuevas fuentes de carbono, la resistencia a antibióticos y metales pesados o la adquisición de patogenicidad (Boyd y col., 2009). En el genoma de TFA se han detectado IGs utilizando el programa IslandViewer 3 (Dhillon y col., 2015) y considerando las predicciones de los métodos SIGI-HMM e IslandPath-DIMOB, los cuales se describen en materiales y métodos. Dichas zonas se indican en el anillo 22 de la Figura 25 (círculo verde). Las principales IGs predichas por ambos métodos se listan en la Tabla 12, mientras que el total de IGs detectadas y su numeración se adjuntan en el Anexo IV.

**Tabla 12. Islas genómicas cuyas coordenadas solapan en las predicciones de los dos métodos utilizados en IslandViewer 3.**

IslandPath-DIMOB				SIGI-HMM				Proteínas
IG	Inicio	Fin	Tamaño (pb)	IG	Inicio	Fin	Tamaño (pb)	
2	1691609	1760539	68930	19	1691609	1722457	30848	Rutas de biodegradación
				21	1734079	1761611	27532	Rutas de biodegradación (Incluye la región similar E)
4	2001817	2012061	10244	24	2000239	2009439	9200	No caracterizadas
6	3066560	3103335	36775	31	3063300	3079834	16534	No caracterizadas, bombas de extrusión de metales pesados, relacionadas con fagos y con transferencia de plásmidos
				32	3093114	3102991	9877	
7	3109158	3161042	51884	33	3107612	3127918	20306	
				34	3128726	3138976	10250	
				35	3141481	3162313	20832	

La mayor IG (~69 kb), numerada como 2 en la predicción de IslandPath-DIMOB y dividida en las IGs 19 y 21 según la predicción de SIGI-HMM (Figura 25, Tabla 12 y Anexo IV), contiene dos agrupaciones de genes, separados por transposasas e integrasas, que codifican para proteínas implicadas en procesos de biodegradación. Otras IGs extensas encontradas por ambos métodos (IGs 6 y 7 de la predicción de IslandPath-DIMOB) codifican proteínas

relacionadas con bombas de metales pesados, conjugación de plásmidos y proteínas relacionadas con fagos. Las IGs desde la 42 a la 47 (ver Anexo IV), predichas solo por el método SIGI-HMM, contienen proteínas implicadas en resistencia a cobre, mercurio y arsénico y diversas bombas de extrusión. Además, estas regiones incluyen genes relacionados con la replicación, partición y conjugación de plásmidos, lo que sugiere que se han incorporado en el genoma de TFA a través de la integración de plásmidos.

Como se refleja en la Figura 25, parte del operón *thn* que contiene los genes de degradación de tetralina está englobado en dos IGs (IGs 29 y 30, predichas por SIGI-HMM). En concreto, se encuentran dentro de estas regiones los genes *thnCA3A4RY* (del SGRAN\_2801 al SGRAN\_2805), *thnB* (SGRAN\_2800) y *thnGHIJKLM* (del SGRAN\_2788 al SGRAN\_2794). A pesar de que los genes *thn* han podido incorporarse en el genoma de TFA por transferencia horizontal, no se ha podido determinar la procedencia del operón mediante la comparación de su secuencia con la base de datos de nucleótidos (nr/nt) que mantiene el NCBI al no encontrarse ninguna secuencia homóloga.

### 1.3.3. Presencia de plásmidos integrados

A pesar de que muchos miembros de la familia *Sphingomonadaceae* contienen grandes plásmidos que contribuyen a su diversidad metabólica y resistencia a metales tóxicos (Stolz, 2014), tanto el ensamblaje del genoma en una única molécula circular como distintos experimentos de lisis y electroforesis en campo pulsante (Hernández Silva, 2000) descartaron la presencia de plásmidos en TFA. Sin embargo, en el proceso de anotación del genoma se identificaron agrupaciones de genes codificantes para proteínas implicadas en la iniciación de la replicación (Rep) y en la partición de los plásmidos (Par) (Figura 25, en azul).

Un posible operón *repAparAG* (SGRAN\_0191, SGRAN\_0190 y SGRAN\_0189) se localiza en la región con homología a *O. carboxidovorans* OM4 (Región 2, ver Figura 25), cerca de un operón *parAB* al que le falta el gen *repA* (SGRAN\_0260 y SGRAN\_0259). Un segundo operón *repAparAG* (SGRAN\_4224, SGRAN\_4225 y SGRAN\_4226) está contenido en la IG 46, detectada por el método SIGI-HMM, muy cerca de las IG 44 y 45, las cuales contienen genes relacionados con la resistencia a metales. El resto de genes *repA* y *parA* se encuentran dispersos por el genoma. Cabe destacar que todas las proteínas RepA y ParA detectadas en el genoma de TFA se clasifican dentro de las super-familias RPA (Pfam 10134) y CbiA (Pfam 01656), respectivamente.

No sólo se han detectado genes relacionados con la replicación de plásmidos sino también un gran número de genes que codifican para proteínas implicadas en el proceso de

## Resultados

conjugación de plásmidos (genes *tra/trb*), distribuidos en 8 agrupaciones (Figura 25, en marrón). Algunas de esas agrupaciones solapan con las zonas de alta similitud con otras bacterias y/o con IGs, lo cual sugiere que los mecanismos de transferencia horizontal han contribuido en gran medida en la incorporación de ADN en el genoma de TFA.

### 1.3.4. Identificación de profagos y secuencias CRISPR

Se han definido tres zonas en el genoma de TFA que podrían corresponder con secuencias de profagos (Figura 25, en amarillo y Tabla 13). Los profagos se detectaron como profagos cuestionables, de acuerdo a la métrica establecida por la herramienta PHAST (*PHAge Search Tool*) (Zhou y col., 2011). Las tres regiones tienen en común genes que codifican para transposasas e integrasas, que constituyen elementos típicos de los profagos.

El profago 1 es el único que contiene genes codificantes para proteínas estructurales de fagos, como son la proteína de la cápsida, proteína portal y proteína de la cola. Sin embargo, solo en el profago 2 se han detectado sitios *att* y un ARN transferente, los cuales suelen ser utilizados por los fagos como sitios de integración. En la Tabla 13 se señalan en negrita las descripciones de los genes con las palabras clave que tiene en cuenta la herramienta PHAST para la predicción de profagos.

Tabla 13. Características de los tres profagos identificados con la herramienta PHAST.

Profago	Inicio	Fin	Tamaño (Kb)	%GC	Genes	Descripción
1	920673	934168	13,4	69,34	SGRAN_0859	<b>Transposasa ISSpo8</b>
					SGRAN_0860	Proteína no caracterizada
					SGRAN_0861	Proteína no caracterizada
					SGRAN_0862	<b>Proteína tipo terminasa</b>
					SGRAN_0863	<b>Proteína portal de fago</b>
					SGRAN_0864	Proteína no caracterizada
					SGRAN_0865	<b>Peptidasa U35 de fago</b>
					SGRAN_0866	<b>Proteína de cápsida de fago</b>
					SGRAN_0867	Peptidasa S74
					SGRAN_0868	Proteína no caracterizada
					SGRAN_0869	Proteína no caracterizada
					SGRAN_0870	Proteína no caracterizada
					SGRAN_0871	Proteína no caracterizada

Profago	Inicio	Fin	Tamaño (Kb)	%GC	Genes	Descripción
2	1681999	1703013	21	62,22	SGRAN_0872	Proteína no caracterizada
					SGRAN_0873	<b>Proteína de la cola del fago</b>
					SGRAN_0874	Proteína no caracterizada
					SGRAN_0875	Proteína no caracterizada
					SGRAN_0876	<b>Proteína que controla la longitud de la cola del fago</b>
					SGRAN_1552	Proteína hipotética
					SGRAN_1553	Subunidad de la glutamato sintasa
					SGRAN_1554	Proteína de resistencia a bacitracina
					SGRAN_1555	Deshidrogenasa/isomerasa
					SGRAN_1556	<b>ARNt-Leu-CAG</b>
					SGRAN_1557	<b>Integrasa</b>
					SGRAN_1558	Proteína no caracterizada
					SGRAN_1559	Nucleasa
					SGRAN_1560	Transglicosilasa lítica
					SGRAN_1561	Proteína no caracterizada
					SGRAN_1562	<b>Regulador transcripcional de fago, AlpA</b>
					SGRAN_1563	Proteína no caracterizada
					SGRAN_1564	<b>ISCC3, transposasa OrfB</b>
					SGRAN_1565	<b>ISCC3, transposasa OrfA</b>
					SGRAN_1566	CoA-ligasa
					SGRAN_1567	<b>Transposasa</b>
					SGRAN_1568	<b>IS511, transposasa OrfB</b>
					SGRAN_1569	Tioesterasa
					SGRAN_1570	Peroxidasa
					SGRAN_1571	Regulador transcripcional MarR
3	4057886	4067397	9,5	67,62	SGRAN_3710	De resistencia a ácido fusárico
					SGRAN_3711	Regulador transcripcional TetR
					SGRAN_3712	Proteína no caracterizada
					SGRAN_3713	Regulador transcripcional XRE
					SGRAN_3714	<b>Transposasa</b>
					SGRAN_3715	<b>Transposasa</b>

## Resultados

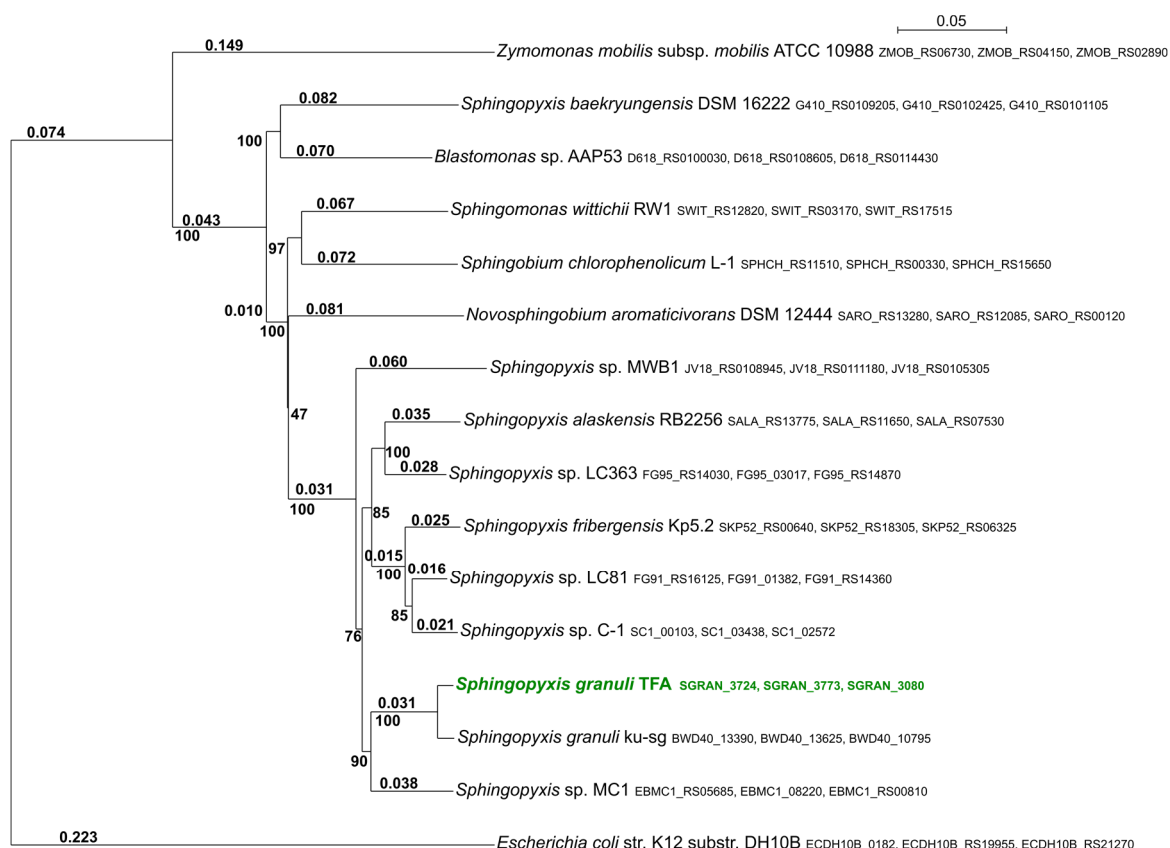
Profago	Inicio	Fin	Tamaño (Kb)	%GC	Genes	Descripción
					SGRAN_3716	Integrasa, dominio unión a ATP
					SGRAN_3717	Integrasa, región catalítica
					SGRAN_3718	Transposasa

Por último, con la aplicación CRISPRfinder (Grissa y col., 2007) se encontraron dos secuencias CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) en el genoma de TFA. La región CRISPR 1 (coordenadas 1982489-1982753), definida como una secuencia CRISPR confirmada, se localiza a unas 300 Kb del profago 2. La región CRISPR 2 es más pequeña y se encuentra entre las bases 4127701 y 4127789 del genoma de TFA. Utilizando este programa no se han detectado proteínas Cas asociadas a los elementos CRISPR. Sin embargo, en el proceso de anotación del genoma con Sma3s, se anotó una proteína Cas (SGRAN\_3113) codificada por un gen exclusivo de TFA, no asociada a las secuencias CRISPR detectadas.

## 2. Genómica comparada

### 2.1. Adscripción filogenética de *S. granuli* TFA al género *Sphingopyxis* mediante árboles filogenéticos

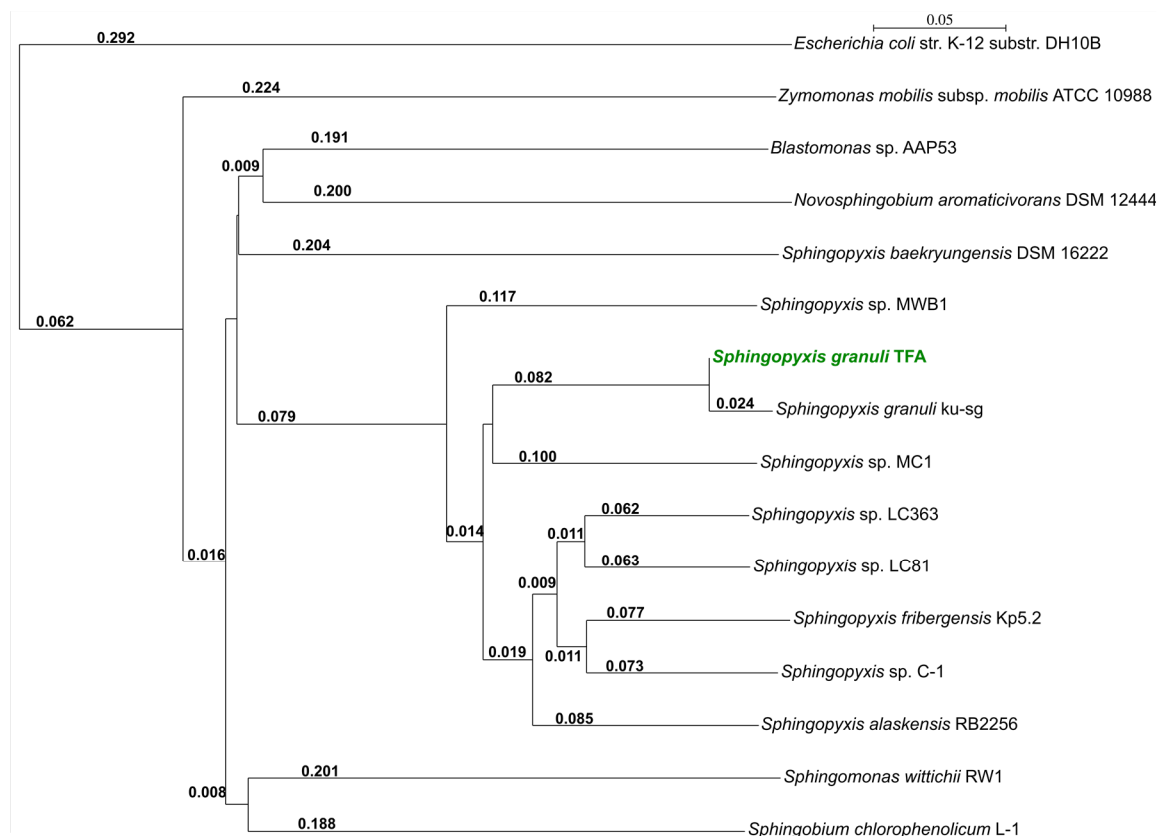
Como se menciona en la introducción, TFA se clasificó inicialmente como una bacteria perteneciente al género *Sphingomonas* por comparación de una secuencia parcial del gen codificante del ARNr 16S. Para contrastar este resultado se han realizado distintos análisis comparativos entre TFA y otros miembros de la familia *Sphingomonadaceae*. En primer lugar, se llevó a cabo un estudio filogenético empleando la secuencia concatenada de los genes codificantes del ARNr 16S (SGRAN\_3724), la subunidad beta de la ATP sintasa (*atpD*; SGRAN\_3773) y la subunidad beta de la polimerasa de ARN (*rpoB*; SGRAN\_3080). Se construyó un árbol filogenético a partir del alineamiento con Clustal Omega de las secuencias de nucleótidos concatenadas de los tres genes anteriores y empleando el método *Neighbor-Joining* (Saitou y Nei, 1987) (Figura 26).



**Figura 26. Análisis filogenético de TFA.** Árbol filogenético construido a partir de las secuencias concatenadas de los genes conservados ARNr 16S, *atpD* y *rpoB*, cuyos *locus\_tag* se indican en el mismo orden tras el nombre de la estirpe. El porcentaje de árboles en el test de bootstrap (1000 réplicas) en el que aparece cada nodo se representa bajo las ramas junto al nodo correspondiente. El número encima de las ramas indica la distancia.

Además, se representó en forma de dendrograma el análisis de la media en identidad de aminoácidos (ver apartado 4.1.2 de materiales y métodos), lo cual proporciona información sobre la relación taxonómica entre organismos a nivel de proteomas completos (Konstantinidis y Tiedje, 2005) (Figura 27).

## Resultados



**Figura 27. Análisis filogenético de TFA.** Dendrograma basado en la media de identidad en aminoácidos (AAI). Encima de las ramas se indica la distancia entre ellas.

TFA se engloba dentro del género *Sphingopyxis*, tanto en el árbol basado en los genes conservados concatenados (ARNr 16S, *atpD* y *rpoB*) (Figura 26) como en el dendrograma de la media en identidad de aminoácidos (Figura 27). Como se observa, el género *Sphingopyxis* forma un grupo compacto del que *Sphingopyxis baekryungensis* DSM 16222 no parece formar parte, por lo que la clasificación actual de esta bacteria debería ser revisada.

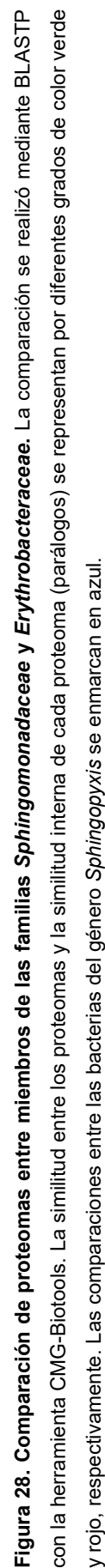


## 2.2. Adscripción filogenética de *S. granuli* TFA al género *Sphingopyxis* mediante una matriz de proteomas

Se compararon los proteomas de distintos miembros pertenecientes a las familias *Sphingomonadaceae* y *Erythrobacteriaceae* a través del programa *matrix* incorporado en el paquete de herramientas CMG-Biotools (Vesth y col., 2013), con el fin de establecer relaciones filogenéticas en base a la similitud entre proteomas. Dicho programa incorpora un algoritmo que permite una comparación a través de BLASTP y por parejas de los distintos proteomas, considerando a dos proteínas ortólogas cuando comparten, al menos, el 30 % de identidad y el porcentaje de secuencia de la proteína más larga cubierto en el alineamiento es igual o mayor al 70 %. El resultado es una matriz triangular donde se estiman cuántas proteínas comparten dos genomas entre sí.

Como se observa en la Figura 28, el rango de similitud entre las bacterias del género *Sphingopyxis* analizadas (excluyendo a *Sphingopyxis baekryungensis*) va desde 45,8 % a 69,6 %, correspondiendo el primer valor a la comparación entre TFA y *Sphingopyxis* sp. C-1 y el segundo a TFA y *Sphingopyxis granuli* ku-sg. Por tanto, TFA, como ya se observó en el dendograma de la Figura 27, presenta un proteoma más similar al de *Sphingopyxis granuli* ku-sg que al resto de *Sphingopyxis* incluidas en el análisis. El mayor porcentaje de similitud entre las bacterias del género *Sphingopyxis* y el resto de bacterias incorporadas al análisis es de 44,2 %, el cual es inferior al menor porcentaje de similitud encontrado entre las bacterias del género *Sphingopyxis*.

Los genomas más grandes dentro del género *Sphingopyxis*, correspondientes a TFA y *S. fribergensis* Kp5.2, son a su vez los que presentan una mayor similitud interna (paralogía), siendo del 13,1 % y 13,4 % respectivamente (intersecciones en rojo de la Figura 28). Esto sugiere que el mayor tamaño de sus genomas no implica una mayor diversidad proteica sino repetición de distintas funciones que han podido ser adquiridas por duplicación y reorganizaciones génicas.



**Figura 28. Comparación de proteomas entre miembros de las familias *Sphingomonadaceae* y *Erythrobacteraceae*.** La comparación se realizó mediante BLASTP con la herramienta CMG-Biotools. La similitud entre los proteomas y la similitud interna de cada proteoma (parálogos) se representan por diferentes grados de color verde y rojo, respectivamente. Las comparaciones entre las bacterias del género *Sphingopyxis* se enmarcan en azul.

### 2.3. Determinación de la especie de *S. granuli* TFA

Para definir la especie de TFA (anteriormente *macrogoltabida*) se calculó la media en identidad de nucleótidos (ANI, *Average Nucleotide Identity*) de los miembros del género *Sphingopyxis* analizados en los árboles filogenéticos y en la matriz de proteomas (Tabla 14). Para ello se compararon las secuencias de los distintos genomas a través de la herramienta ANIm, disponible en servidor JSpeciesWS (Richter y col., 2015). En dicha comparación, los genomas más parecidos entre sí son el de TFA y *S. granuli* ku-sg, con un 97,76 % de media en identidad de nucleótidos y cubriendo en el alineamiento el 72,45 % y 71,62 % de sus genomas, respectivamente. Por otra parte, los alineamientos del genoma de *S. baekryungensis* DSM 16222 con el resto de bacterias del género *Sphingopyxis* presentan una baja cobertura de secuencia, que en ninguno de los casos alcanza un 5 %.

**Tabla 14. Resultado del análisis de ANIm en porcentaje de identidad y, entre paréntesis, porcentaje de secuencia incluida en el alineamiento.** En verde se muestran las comparaciones con un valor por encima del 95 % y en rojo alineamientos con baja cobertura de secuencia. Los números del 1 al 10 hacen referencia a (1) *Sphingopyxis alaskensis* RB2256, (2) *Sphingopyxis granuli* TFA, (3) *Sphingopyxis* sp. C-1, (4) *Sphingopyxis* sp. LC363, (5) *Sphingopyxis* sp. LC81, (6) *Sphingopyxis fribergensis* Kp5.2, (7) *Sphingopyxis* sp. MC1, (8) *Sphingopyxis granuli* ku-sg, (9) *Sphingopyxis* sp. MWB1, y (10) *Sphingopyxis baekryungensis* DSM 16222.

	1	2	3	4	5	6	7	8	9	10
1	*	85,50 (48,00)	85,43 (56,33)	86,78 (64,51)	86,06 (62,42)	85,83 (60,12)	85,86 (49,32)	85,29 (45,55)	84,51 (30,32)	<b>82,60</b> <b>(3,05)</b>
2	85,50 (34,70)	*	84,58 (32,88)	85,54 (39,84)	85,13 (38,93)	84,97 (36,26)	85,68 (39,42)	<b>97,76</b> <b>(71,62)</b>	84,45 (20,27)	<b>83,09</b> <b>(2,91)</b>
3	85,44 (41,48)	84,57 (33,43)	*	86,63 (56,29)	87,99 (64,45)	87,10 (61,79)	84,97 (33,92)	84,69 (32,62)	<b>84,02</b> <b>(18,73)</b>	<b>82,88</b> <b>(2,34)</b>
4	86,76 (52,40)	85,54 (44,41)	86,63 (62,07)	*	87,77 (69,41)	88,03 (70,87)	85,78 (43,00)	85,35 (41,37)	84,40 (24,04)	<b>82,73</b> <b>(2,32)</b>
5	86,07 (47,88)	85,13 (41,12)	88,00 (67,36)	87,77 (65,72)	*	88,83 (67,89)	85,13 (38,74)	85,06 (38,64)	84,25 (22,36)	<b>82,92</b> <b>(2,38)</b>
6	85,83 (38,98)	84,97 (32,44)	87,10 (54,57)	88,03 (57,00)	88,83 (57,76)	*	85,22 (32,74)	84,71 (29,94)	<b>84,22</b> <b>(17,06)</b>	<b>82,88</b> <b>(1,97)</b>
7	85,86 (45,18)	85,69 (50,07)	84,98 (42,54)	85,77 (48,78)	85,13 (46,41)	85,22 (46,34)	*	85,58 (47,60)	84,56 (25,99)	<b>82,87</b> <b>(2,89)</b>
8	85,29 (33,35)	<b>97,76</b> <b>(72,45)</b>	84,69 (32,63)	85,35 (37,44)	85,06 (36,90)	84,71 (33,84)	85,57 (37,91)	*	84,46 (20,23)	<b>82,70</b> <b>(2,77)</b>
9	84,50 (32,76)	84,46 (30,26)	84,01 (27,56)	84,40 (32,02)	84,25 (31,42)	84,23 (28,12)	84,58 (30,40)	84,46 (29,74)	*	<b>82,71</b> <b>(3,12)</b>
10	<b>82,60</b> <b>(3,47)</b>	<b>83,09</b> <b>(4,52)</b>	<b>82,88</b> <b>(3,63)</b>	<b>82,73</b> <b>(3,30)</b>	<b>82,92</b> <b>(3,54)</b>	<b>82,88</b> <b>(3,39)</b>	<b>82,87</b> <b>(3,54)</b>	<b>82,70</b> <b>(4,24)</b>	<b>82,71</b> <b>(3,26)</b>	*

## Resultados

Hasta el momento, en el análisis de filogenia se han utilizado genomas secuenciados completamente o divididos en fragmentos. Para asignar una especie a TFA considerando un mayor número de bacterias, se realizó un BLASTN del gen codificante del ARNr 16S de TFA (SGRAN\_3724) a través del servidor EzBioCloud (<http://www.ezbiocloud.net/>) (Yoon y col., 2017), que incorpora una amplia base de datos de secuencias del gen codificante del ARNr 16S.

Los resultados obtenidos (Tabla 15) mostraron que la especie más similar a TFA es *Sphingopyxis granuli* Kw07(T) con una similitud del 99,65 %. Cabe indicar que, a la fecha de realización del estudio, la especie *S. granuli* ku-sg no estaba incorporada en dicha base de datos.

Tabla 15. Resultado del alineamiento del gen codificante del ARNr 16S en el servidor EzBioCloud.

Nombre	Estirpe	Nº de acceso	Similitud (%)	Diferente/ Total nt	Completo (%)
<i>Sphingopyxis granuli</i>	Kw07(T)	AY563034	99,65	5/1422	100
<i>Sphingopyxis ginsengisoli</i>	Gsoil250(T)	AB245343	98,34	24/1442	100
<i>Sphingopyxis ummariensis</i>	UI2(T)	EF424391	98,01	28/1408	99,51
<i>Sphingopyxis panaciterrulae</i>	DCY34(T)	EU075217	97,97	29/1432	100
<i>Sphingopyxis witflariensis</i>	W-50(T)	AJ416410	97,79	32/1447	100
<i>Sphingopyxis panaciterrae</i>	Gsoil124(T)	AB245353	97,78	32/1444	100
<i>Sphingopyxis italica</i>	SC13E-S71(T)	HE648058	97,73	32/1411	99,86
<i>Sphingopyxis indica</i>	D515(T)	EF494193	97,72	32/1401	99,08
<i>Sphingopyxis macrogoltabida</i>	IFO15033(T)	D13723	97,59	34/1410	99,02

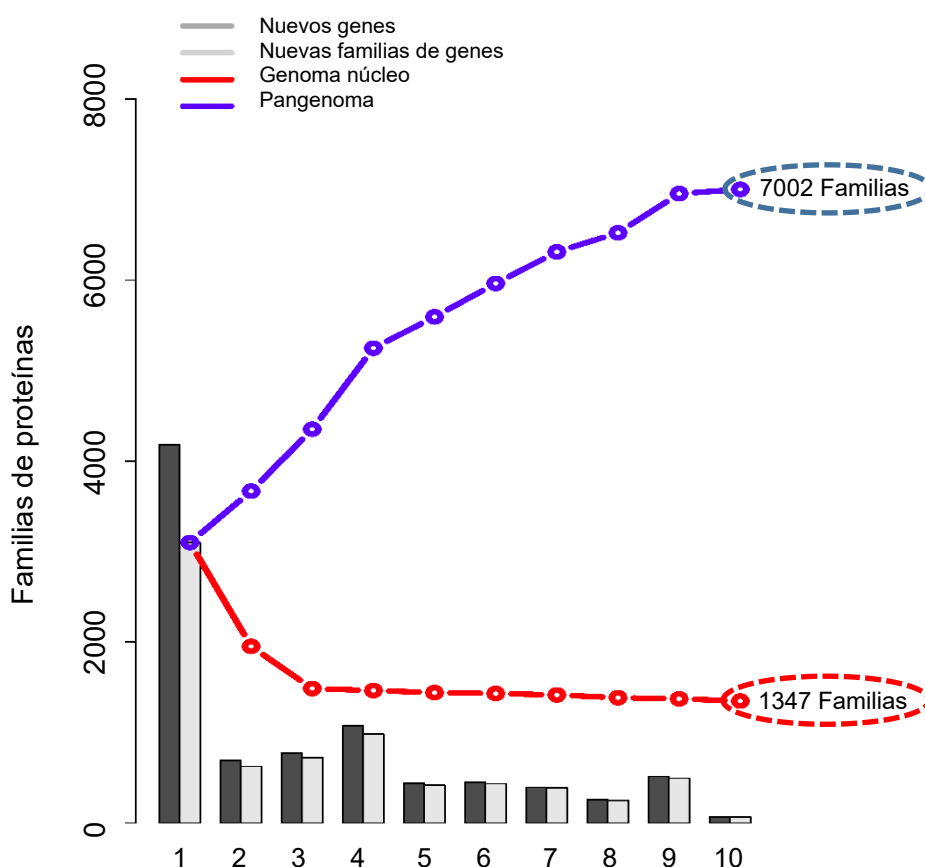
De acuerdo a este resultado, a la cercanía con *Sphingopyxis granuli* ku-sg en los árboles filogenéticos (Figura 26 y Figura 27), a la similitud determinada a nivel de proteínas en la matriz de proteomas (Figura 28) y al resultado del análisis de ANIm (Tabla 14), TFA pasó a clasificarse como *Sphingopyxis granuli* estirpe TFA.

### 2.4. Análisis del pangenoma y el genoma núcleo del género *Sphingopyxis*

Para determinar las proteínas exclusivas de TFA, se analizó el pangenoma y el genoma núcleo de las bacterias del género *Sphingopyxis* examinadas en la matriz de proteomas y en los árboles filogenéticos, utilizando para ello el programa *pancoreplot* incluido en la plataforma GMC-Biotools. El genoma núcleo de un conjunto de bacterias lo constituyen las familias de

proteínas comunes en todas ellas, mientras que el pangenoma representa de forma no redundante el genoma núcleo más el resto de familias de proteínas contenidas en los distintos genomas.

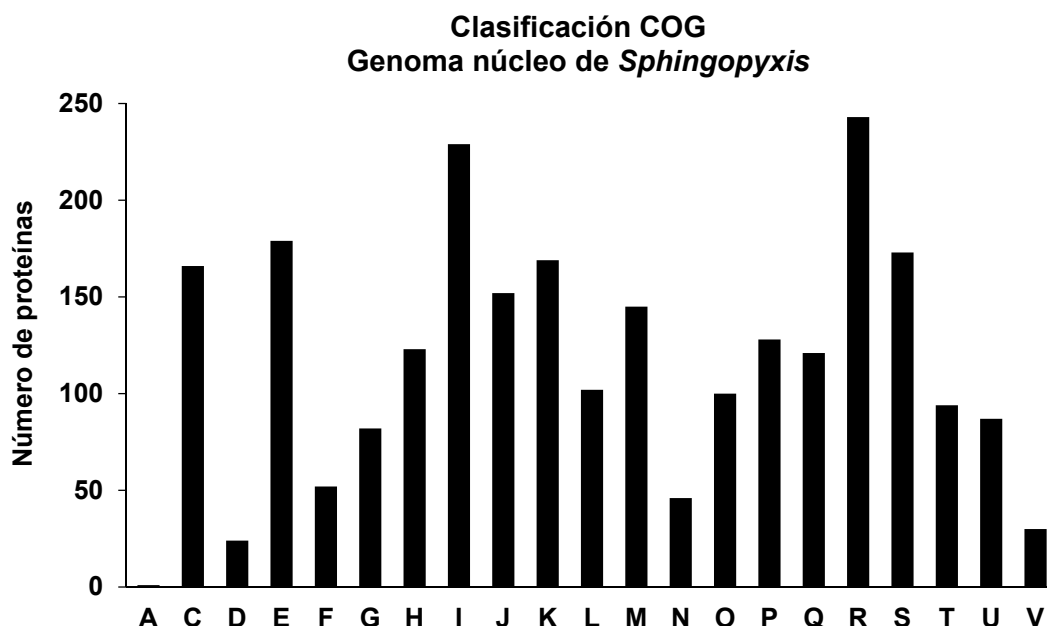
La Figura 29 muestra la evolución del número de familias de proteínas en el pangenoma y el genoma núcleo a medida que se incorporan las distintas *Sphingopyxis* al análisis. El número de familias de proteínas comunes a todas las bacterias es de 1347 (genoma núcleo) y la diversidad de proteínas en el pangenoma es de 7002 familias. Las 1347 familias del genoma núcleo incluyen 2271 proteínas diferentes de TFA, lo cual representa más de la mitad (54,2 %) de los genes codificantes para proteínas anotados en TFA.



**Figura 29. Representación del pangenoma y el genoma núcleo del género *Sphingopyxis*.** Los números del eje X hacen referencia a las distintas bacterias incluidas en el análisis: (1) *S. granuli* TFA, (2) *S. alaskensis* RB2256, (3) *S. baekryungensis* DSM 16222, (4) *Sphingopyxis* sp. Kp5.2, (5) *Sphingopyxis* sp. LC363, (6) *Sphingopyxis* sp. LC81, (7) *Sphingopyxis* sp. MC1, (8) *Sphingopyxis* sp. MWB1, (9) *Sphingopyxis* sp. C-1 y (10) *S. granuli* ku-sg. Rodeados con línea discontinua se señalan el número de familias en el genoma núcleo (rojo) y el pangenoma (azul).

## Resultados

Los genes de TFA incluidos en el genoma núcleo final del género *Sphingopyxis* se clasificaron en base a sus categorías COG (*Clusters of Orthologous Groups*) (Figura 30).



**Figura 30. Clasificación de las proteínas pertenecientes al genoma núcleo del género *Sphingopyxis*.** Las categorías funcionales son: (A) procesamiento y modificación del ARN, (C) producción y conversión de energía, (D) control del ciclo celular, (E) transporte y metabolismo de aminoácidos, (F) transporte y metabolismo de nucleótidos, (G) transporte y metabolismo de carbohidratos, (H) transporte y metabolismo de coenzimas, (I) transporte y metabolismo de lípidos, (J) traducción, estructura ribosómico y biogénesis, (K) transcripción, (L) replicación, recombinación y reparación, (M) biogénesis de la envuelta y la membrana celular, (N) movilidad celular, (O) modificación postranscripcional, síntesis y degradación de proteínas y chaperonas, (P) transporte y metabolismo de iones inorgánicos, (Q) catabolismo, transporte y biosíntesis de metabolitos secundarios, (R) predicción de función general, (S) función desconocida, (T) mecanismos de transducción de señal, (U) tráfico intracelular, secreción y transporte vesicular y (V) mecanismos de defensa.

Las categorías más representadas, de mayor a menor número de proteínas, son la de predicción de función general (categoría R), el transporte y el metabolismo de lípidos (categoría I), el transporte y metabolismo de aminoácidos (categoría E) y proteínas de función desconocida (categoría S). Por el contrario, las categorías de procesamiento y modificación del ARN (A) y de control del ciclo celular (D) son las que contienen un menor número de proteínas. Debido a que las categorías A y D están altamente conservadas en los genomas bacterianos, el hecho de que existan pocas proteínas de estos grupos en el genoma núcleo no refleja que la mayoría estén en el pangenoma dispensable, sino que se trata de categorías con pocas proteínas. Además, las categorías de movilidad celular (N) y de mecanismos de defensa (V) también contienen una baja representación en el genoma núcleo.

En cuanto a las funciones específicas, dentro del genoma núcleo están presentes proteínas implicadas en resistencia a cobre y bombas de extrusión de metales pesados, además de un gran número de acil-CoA deshidrogenasas de ácidos grasos (54), proteínas ribosómicas (56), transportadores de tipo ABC (22), proteínas relacionadas con la síntesis y movilidad del flagelo (21) y otras proteínas de función conservada como enzimas implicadas en el metabolismo central o polimerasas de ARN y ADN.

En relación con las proteínas Thn, las codificadas por *thnA1A2A3A4*, que constituyen los componentes de la dioxigenasa inicial de la ruta, así como las codificadas por *thnCDEFM* y el coactivador ThnY, se encuentran fuera del genoma núcleo, mientras que las enzimas implicadas en la  $\beta$ -oxidación del ácido pimélico (intermediario de la degradación de THN), el regulador ThnR, la dioxigenasa ThnB y la aldehído deshidrogenasa ThnG se engloban dentro del genoma núcleo.

La generación del pangenoma y el genoma núcleo del género *Sphingopyxis* permitió la identificación de las proteínas específicas de TFA (ver Anexo V), las cuales se obtuvieron ejecutando el programa *pancoreplot\_subsets* con las opciones -o 1 -i 1 -c 2:10 en la herramienta CMG-Biotools. En el análisis se obtuvieron 303 proteínas exclusivas de TFA, dentro de las cuales se encuentran 128 proteínas no caracterizadas, una deshidrogenasa de alcohol de polivinilo (PVA) y una hidrolasa de alcohol de polivinilo oxidado, ambas implicadas en la degradación de este compuesto, y numerosas transposasas, integrasas y proteínas de transferencia y conjugación de plásmidos. Además, las proteínas codificadas por el operón *narUGHJI*, responsables de la respiración anaerobia de nitrato (García-Romero y col., 2016), y el gen *ptrB* divergente a dicho operón sólo se encuentran en las estirpes analizadas de la especie *granuli*.

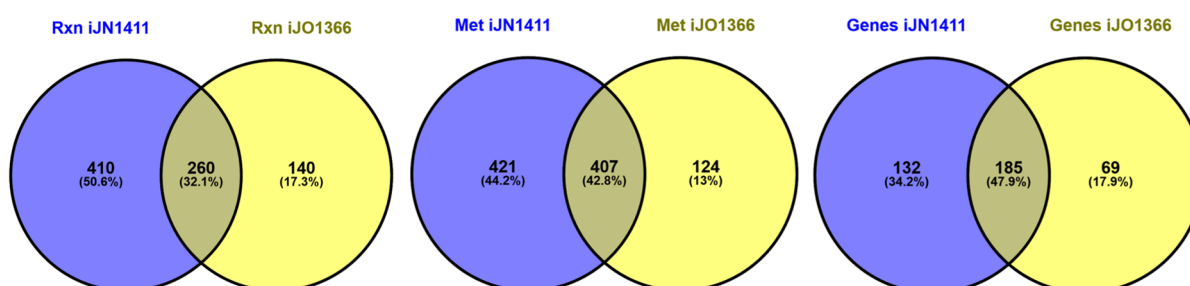
### 3. Reconstrucción metabólica de *S. granuli* TFA

La reconstrucción metabólica de *S. granuli* TFA, se realizó inicialmente a partir de la anotación funcional del genoma que se describe en esta Tesis doctoral. En primer lugar se identificaron proteínas ortólogas entre TFA y *Pseudomonas putida* KT2440 y entre TFA y *Escherichia coli* K12. Las listas de ortólogos obtenidas se emplearon posteriormente para generar un primer borrador del modelo metabólico de TFA a través de la asignación automática de reacciones bioquímicas a los genes de TFA según las asignadas a sus genes ortólogos en *P. putida* KT2440 y *E. coli* K12 en los modelos metabólicos *iJN1411* (Nogales y col., 2017) y *iJO1366* (Orth y col., 2011), respectivamente. Esta primera versión del modelo contenía 810 reacciones, 952 metabolitos y 386 genes (Figura 31). De las reacciones y metabolitos asignados a este primer borrador, 260 reacciones y 407 metabolitos fueron

## Resultados

añadidos procedentes tanto del modelo *iJN1411* como del modelo *iJO1366*, conformando el conjunto de reacciones y metabolitos comunes a dichos modelos y al borrador del modelo de TFA. Por otro lado, 410 reacciones y 421 metabolitos fueron asignados exclusivamente desde el modelo *iJN1411*, mientras que solo 140 reacciones y 124 metabolitos se incorporaron a partir del modelo *iJO1366*, lo cual sugiere que TFA es metabólicamente más parecida a *P. putida* que a *E. coli*.

En este modelo inicial, se incluyeron 386 de los 4190 genes anotados en TFA que codifican para proteínas, lo cual supone un 9,2 % del total de proteínas, un porcentaje muy bajo incluso si en el cálculo se excluyen las proteínas no metabólicas.



**Figura 31. Representación con diagramas de Venn de las reacciones (Rxn), metabolitos (Met) y genes (Genes) incorporados en el borrador inicial de modelo de TFA.** En azul se representan los elementos incorporados al modelo a partir del modelo *iJN1411* (*P. putida* KT2440) y en amarillo los procedentes del modelo *iJO1366* (*E. coli* K12).

Un segundo borrador del modelo de TFA, que se obtuvo mediante la herramienta de anotación SEED, consistió en 1074 reacciones, 992 metabolitos y 4428 genes. Dichos genes fueron anotados de forma automática por la herramienta a partir de la secuencia genómica de TFA y, en base a dicha anotación, la propia herramienta generó un nuevo borrador del modelo de TFA.

Ambos borradores, el basado en los modelos *iJN1411* e *iJO1366* y el generado por la herramienta SEED, se fusionaron en uno solo eliminando las reacciones y los metabolitos repetidos. A continuación, se validaron una por una todas las reacciones incluidas en el borrador como se indica en la apartado 3.2.2 de materiales y métodos y, según el resultado, se mantuvieron o se excluyeron del modelo. Atendiendo a los rasgos fenotípicos descritos para la familia *Sphingomonadaceae* (Glaeser y Kämpfer, 2014), a la que pertenece TFA, se incorporaron las reacciones de síntesis de esfingolípidos, hasta el precursor esfingánina (dihidroesfingosina). Para obtener dicho precursor se incorporó en el modelo una reacción llevada a cabo por la enzima serina palmitoiltransferasa (SPT) que cataliza la primera reacción para la síntesis *de novo* de esfingolípidos, como se describió para *Sphingomonas wittichii*



RW1 (Raman y col., 2010). La enzima de TFA, codificada por el gen SGRAN\_1009, presenta una identidad del 76,2 % con la SPT de *Sphingomonas wittichii* RW1, codificada por el gen Swit\_3900, con un 99,75 % de cobertura de la secuencia. La reacción que lleva a cabo la enzima SPT consiste en una condensación descarboxilativa tipo Claisen de la L-serina y el palmitoil-CoA para generar 3-keto-dihidroesfingosina (3- dehidroesfinganina), un producto de cadena larga que se transforma en esfinganina mediante la enzima 3-dehidroesfinganina reductasa (no encontrada en TFA). La esfinganina, como precursor de esfingolípidos, se incorporó a la reacción de biomasa.

En relación con las características metabólicas de TFA, se introdujeron en el modelo las reacciones para la degradación de la tetralina, descritas en la introducción de esta Tesis, para la respiración anaerobia de nitrato, para la síntesis de polihidroxibutirato (PHB) y para la síntesis de ectoína. En trabajos anteriores se había medido la producción de PHB (Martín-Cabello y col., 2011) y la capacidad de TFA para crecer anaeróbicamente con nitrato (García-Romero y col., 2016), confirmando experimentalmente que ambos procesos ocurren en la bacteria. Sin embargo, la producción de ectoína sólo se ha inferido de los datos de la anotación funcional sin que se haya validado experimentalmente su producción en TFA.

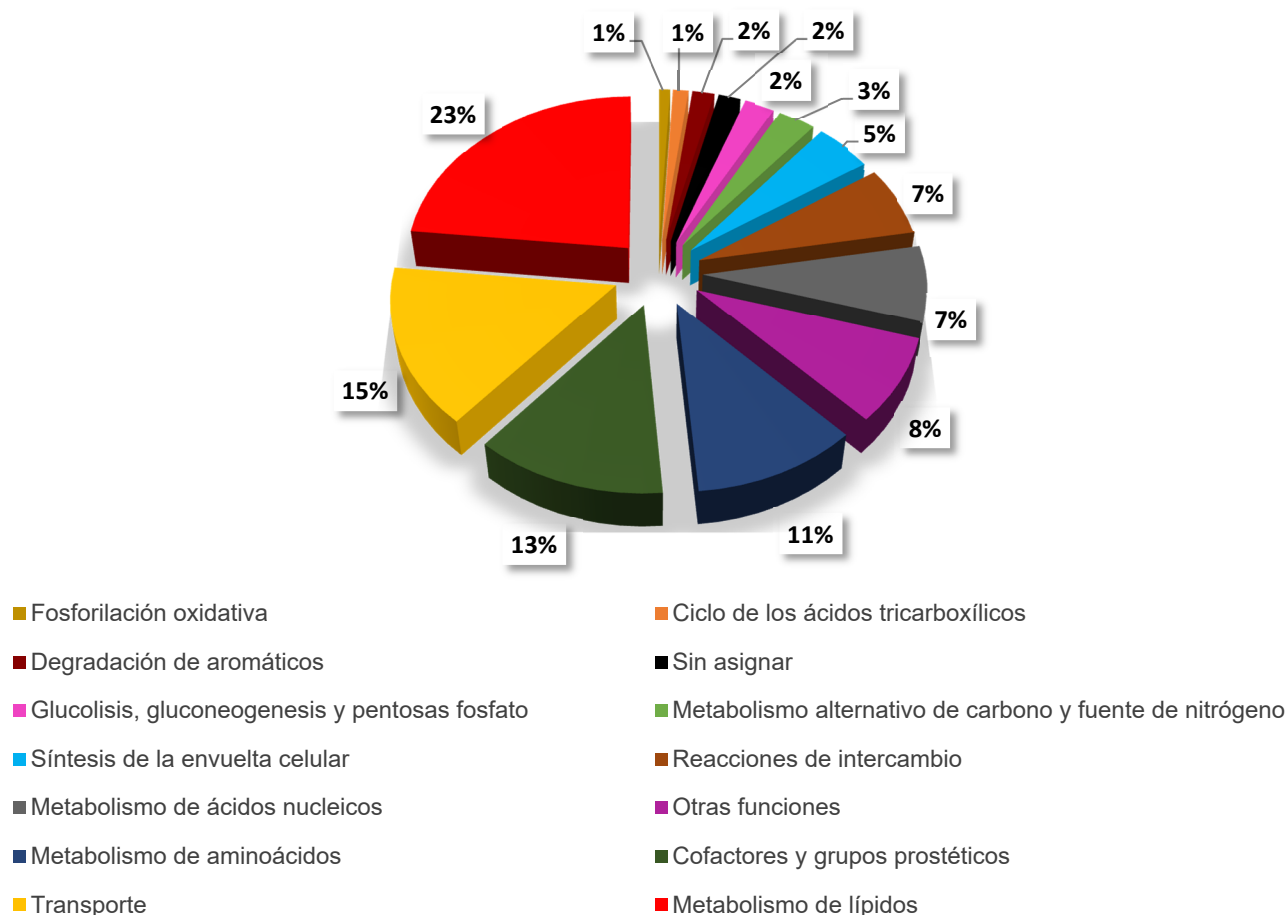
La ectoína es un soluto compatible que ayuda a la supervivencia de los microorganismos en condiciones de estrés osmótico, actuando como un osmolito. Se encuentra en alta concentración en algunas bacterias halófilas para protegerlas frente a la salinidad (Roberts, 2005). El operón que codifica para las enzimas implicadas en la síntesis de ectoína y 5-hidroxiectoína en TFA, *ectABCD*, lo conforman los genes SGRAN\_2345, SGRAN\_2346, SGRAN\_2347 y SGRAN\_2348, respectivamente. Para modelar la ruta de síntesis de la ectoína se utilizó la información disponible en MetaCyc y para representar en el modelo la hidroxilación de la ectoína y generación de 5-hidroxiectoína, catalizada por EctD, se añadió la reacción descrita por Reuter y col. (2010).

En cuanto al metabolismo de PHB, a pesar de tener evidencias experimentales de su producción, no se han estudiado ni las rutas de biosíntesis ni de despolimerización en TFA, por lo que se asignaron las reacciones descritas para *E. coli* (Wang y col., 2009). Además, se incorporó una reacción de demanda de PHB (denominada DM\_PHB) para poder introducir en el modelo un flujo de producción de este polímero cuando fuese necesario.

Finalmente, se añadieron 85 reacciones de intercambio de metabolitos con el medio y otras reacciones no asociadas a genes requeridas para el funcionamiento del modelo (ver apartado 3.2.2 de materiales y métodos).

### 3.1. Descripción general del modelo metabólico de *S. granuli* TFA

El modelo final obtenido, *ilG738*, consta de 1389 reacciones, 1107 metabolitos y 738 genes. Todas las reacciones se dividen en distintos subsistemas atendiendo a la ruta metabólica a la que pertenecen (Figura 32).



**Figura 32. Clasificación de las reacciones del modelo *ilG738*.** Las reacciones se agruparon en el modelo en 115 subsistemas, agrupándose aquí en 14 para facilitar el análisis.

En TFA, el subsistema que presenta un mayor número de reacciones es el de metabolismo de lípidos. El siguiente subsistema más representado lo constituyen las reacciones de transporte, lo cual no debe extrapolarse a un porcentaje alto de metabolitos transportados, puesto que para la mayoría de ellos se representan en el modelo dos reacciones de transporte, una a través de la membrana externa y otra a través de la interna. A continuación encontramos la síntesis de grupos prostéticos y cofactores y reacciones relacionadas con el metabolismo de aminoácidos.

En cuanto a la degradación de compuestos aromáticos, solo el 2 % de las reacciones del modelo están relacionadas con este subsistema. Entre ellas encontramos las reacciones de degradación de la tetralina, de catabolismo de triptófano, tirosina y fenilalanina y de degradación de fenilpropanoato. Cabe destacar que es la primera vez que se incorpora a un modelo metabólico la ruta aerobia de degradación de tetralina, debido a que TFA es la única bacteria donde se ha caracterizado completamente (ver apartado 3 de la introducción).

El modelo metabólico *i*IG738 contiene las rutas completas para el anabolismo de las macromoléculas necesarias para el crecimiento de TFA como son las de todos los aminoácidos, ácidos nucleicos, lípidos, cofactores y azúcares. Además, el metabolismo respiratorio de TFA se ha definido de forma detallada en el modelo, tanto con la inclusión de las reacciones de fosforilación oxidativa como con las necesarias para la utilización del oxígeno (respiración aerobia) o del nitrato (respiración anaerobia) como último aceptor de electrones.

Los modelos publicados hasta la fecha de bacterias pertenecientes a la clase  $\alpha$ -proteobacteria se listan en la Tabla 16. Dentro de esta categoría filogenética sólo se han generado modelos de 7 especies distintas, de los cuales, el de *Zymomonas mobilis* y el descrito en esta Tesis constituyen los únicos dentro del orden *Sphingomonadales*.

El modelo *i*EM439 de *Zymomonas mobilis* y el modelo *i*Rsp1095 de *Rhodobacter sphaeroides* contienen el mayor porcentaje de proteínas con respecto al conjunto de proteínas codificadas en sus genomas, un 24 % y 23,8 % respectivamente. Valorando este parámetro, TFA se encuentra en el cuarto lugar, con un 17,6 % de su proteoma representado en el modelo *i*IG738.

Atendiendo al número de reacciones y metabolitos, el modelo *i*IG738 es el que presenta un mayor número, 1389 y 1107 respectivamente, lo que indica que es el modelo de la clase  $\alpha$ -proteobacteria con una mayor complejidad metabólica.

Tabla 16. Comparación de los distintos modelos metabólicos publicados de bacterias miembros de la clase  $\alpha$ -proteobacteria.

Características	<i>/EM439</i> <i>Z. mobilis</i>	<i>/Rsp1095</i> <i>R. sphaeroides</i>	<i>/HZ565</i> <i>S. melloti</i>	<i>/OR450</i> <i>R. etli</i>	<i>/RP911</i> <i>M. extorquens</i>	<i>/WZ663</i> <i>K. vulgare</i>	<i>/XW433</i> <i>G. oxydans</i>	<i>/IG738</i> <i>S. granuli</i>
<b>Genes</b>								
codificantes para proteínas en el genoma	1823 <sup>a</sup>	4410 <sup>a</sup>	6218 <sup>a</sup>	5973 <sup>a</sup>	6199 <sup>a</sup>	3054 <sup>a</sup>	2607 <sup>a</sup>	4190
Genes (% del genoma)	439 (24%)	1049 (23,8%) 46 ARNt	565 (9%)	450 (7,5%)	911 (14,7%)	663 (21,7%)	433 (16,6%)	738 (17,6%)
Total de reacciones	692	1158	503	402	1139	830	859	1389
Reacciones con gen asociado	585	1049	481	339	936	621	752 <sup>b</sup>	1057
Reacciones sin gen asociado	107 (15,5%)	109 (9,4%)	22 (4,4%)	63 (15,7%)	203 (17,8%)	209 (25%)	107 (12,5%)	332 (23,9%)
Metabolitos	658	1096	522	377	977	649	985	1107
Referencia	Motamedian y col., 2016	Imam y col., 2011	Zhao y col., 2012	Resendis-Antonio y col., 2012	Peyraud y col., 2011	Zou y col., 2012	Wu y col., 2014	Esta tesis (No publicado)

<sup>a</sup> Información obtenida de la base de datos Genome List (<https://www.ncbi.nlm.nih.gov/genome/browse/>).

### 3.2. Validación del modelo metabólico *ilG738* de *S. granuli* TFA

#### 3.2.1. Validación cualitativa de la utilización de fuentes de carbono

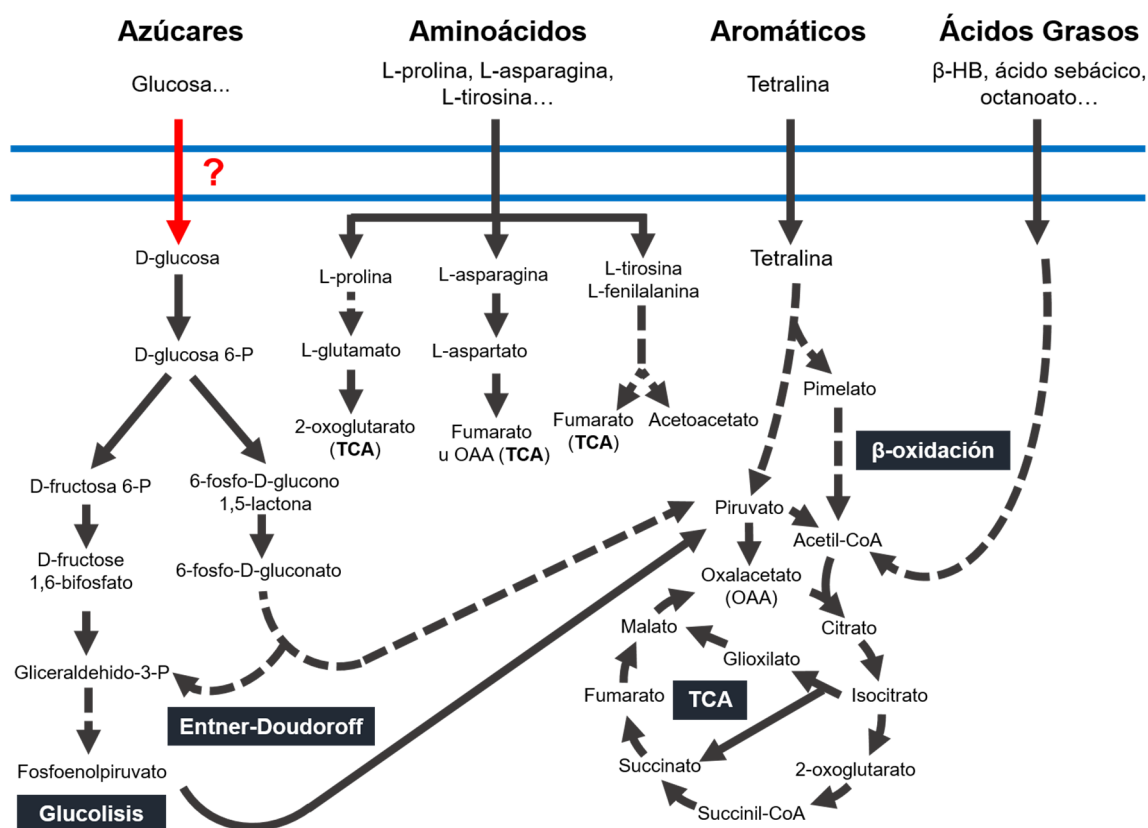
Para evaluar la capacidad de predicción del modelo *ilG738*, se simuló el crecimiento de TFA en un medio mínimo definido con cada una de las distintas fuentes de carbono en las que habitualmente se cultiva a TFA, como son la tetralina, el ácido sebácico y el  $\beta$ -hidroxibutirato. Además, se incorporaron al análisis otras fuentes de carbono en las que TFA podría crecer según la información obtenida en la reconstrucción (Tabla 17). La capacidad de *ilG738* para predecir el crecimiento utilizando distintas fuentes de carbono indicaría que se trata de un modelo robusto, altamente conectado y sin huecos en las rutas metabólicas para el catabolismo de dichos metabolitos. Según los resultados expuestos en la Tabla 17, el modelo predice crecimiento acorde a lo observado *in vivo* en tetralina, ácido sebácico,  $\beta$ -hidroxibutirato, L-lactato, L-prolina, ácido oleico, octanoato, butanoato, ácido pimélico, L-glutamina, L-tirosina, L-triptófano, L-fenilalanina, L-asparagina y acetato. Sin embargo, el modelo predice crecimiento en otras fuentes de carbono donde TFA no mostró crecimiento *in vivo*, como son etanol, formaldehído, 3-fenilpropanoato y 3-(3-hidroxi-fenil)propanoato.

**Tabla 17. Lista de metabolitos testados en medio mínimo como única fuente de carbono para TFA.** La predicción del modelo *ilG738* se muestra en la columna *in silico* y la comprobación experimental en la columna *in vivo*. Verde denota crecimiento, rojo no crecimiento y blanco no analizado.

Metabolito	<i>In silico</i>	<i>In vivo</i>	Metabolito	<i>In silico</i>	<i>In vivo</i>
Etanol			3-fenilpropanoato		
(R,R)-2,3-Butanediol			L-lactato		
Ácido oleico			Dipeptido ala-his		
Glicerol			2-Dehidro-D-gluconato		
Octanoato			3-(3-hidroxi-fenil)propanoato		
Ácido sebácico			Agmatina		
Butanoato			L-histidina		
Dipeptido ala-leu			Dipeptido ala-trp		
L-prolina			D-tyrosine		
Putrescina			L-tirosina		
$\beta$ -hidroxibutirato			Acetato		
Manosa			2,5-diketo-D-gluconato		
Fructosa			L-glutamina		
Glucosa			L-triptófano		
Ácido pimélico			L-asparagina		
Tetralina			L-aspartato		
D-alanina			L-fenilalanina		
Dipeptido ala-lhr			D-fenilalanina		
Dipeptido bala-bala			Coniferol		
Formaldehído					

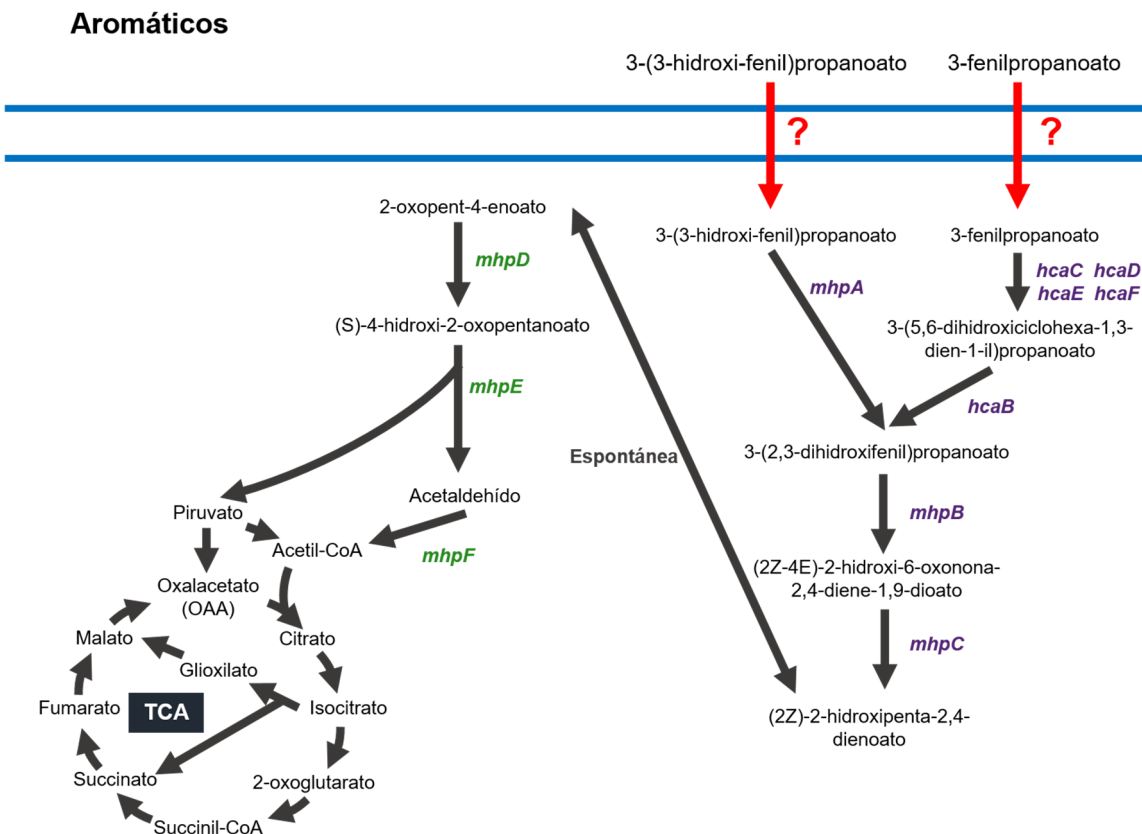
## Resultados

Estos resultados sugieren que TFA presenta un metabolismo adaptado principalmente a la utilización de aminoácidos y ácidos grasos como fuente de carbono y, a pesar de la presencia de numerosas dioxigenasas y otras enzimas relacionadas con el metabolismo de compuestos aromáticos, TFA sólo mostró crecimiento en tetralina y en los tres aminoácidos aromáticos (fenilalanina, tirosina y triptófano). En cuanto al uso de azúcares, TFA no es capaz de crecer ni en glucosa ni en fructosa como única fuente de carbono y energía, a pesar de que en la anotación del genoma se han podido identificar todas las enzimas implicadas en la glucólisis y además la ruta de Entner-Doudoroff (Figura 33).



**Figura 33. Representación esquemática de las principales rutas catabólicas incluidas en el modelo iG738.** Las líneas discontinuas indican conexiones entre metabolitos en las que intervienen más de una reacción. La línea roja denota que no hay evidencias claras, ni experimentales ni por similitud con proteínas homólogas, para incluir la reacción en el modelo. Las líneas azules representan las membranas interna y externa.

A pesar de no observar crecimiento en los metabolitos aromáticos 3-fenilpropanoato y 3-(3-hidroxi-fenil)propanoato, la ruta completa de degradación de estos compuesto pudo reconstruirse en el modelo iIG738 (Figura 34).



**Figura 34. Representación esquemática de las rutas de degradación de 3-(3-hidroxi-fenil)propanoato y 3-fenilpropanoato según la información disponible en la base de datos MetaCyc.** Las líneas rojas denotan que no se han encontrado genes codificantes para transportadores de dichos compuestos en el genoma de TFA y que no se ha observado crecimiento de TFA usándolos como únicas fuentes de carbono y energía. Los genes indicados en verde forman parte de la ruta baja de la degradación, común a otros compuestos, y los morados a la ruta alta, específicos de la degradación de estos metabolitos.

En la anotación del genoma de TFA se identificaron dos grupos de genes que codificaban para las tres proteínas implicadas en la ruta baja de la degradación de 3-fenilpropanoato y 3-(3-hidroxi-fenil)propanoato (señaladas en verde en la Figura 34), así como las enzimas HcaE y HcaF. Para asignar el resto de proteínas de la ruta se realizaron BLASTP frente al proteoma de TFA de las proteínas que llevan a cabo las distintas reacciones en otras bacterias (Tabla 18).

## Resultados

**Tabla 18. Proteínas implicadas en la degradación de 3-fenilpropanoato y 3-(3-hidroxi-fenil)propanoato en TFA.**

Proteína	Descripción	Posibles gen/es codificante en TFA	Mejor resultado en Swiss-Prot	Identidad (%)
MhpD	2-ceto-4-pentanoato hidratasa	SGRAN_1420	MhpD (A6TAC6, <i>Klebsiella pneumoniae</i> )	52,4
		SGRAN_1581	MhpD (Q13VUO, <i>Paraburkholderia xenovorans</i> , tercer resultado)	48,8
MhpF	Acetaldehído deshidrogenasa	SGRAN_1421	Swit_4924 (A5VGU5, Acetaldehído deshidrogenasa de <i>Sphingomonas wittichii</i> )	81,4
		SGRAN_1580	MhpF (A1K6Y5, <i>Azoarcus</i> sp. (strain BH72))	78,2
MhpE	4-hidroxi-2-oxovalerato aldolasa	SGRAN_1422	Swit_4923 (A5VGU4, 4-hidroxi-2-oxovalerato aldolasa de <i>Sphingomonas wittichii</i> )	88,6
		SGRAN_1579	NahM (P51017, 4-hidroxi-2-oxovalerato aldolasa de <i>Pseudomonas putida</i> )	65,6
MhpA	3-(3-hidroxi-fenil)propionato hidroxilasa	SGRAN_3916	Sin similitud a MhpA. Se parece a una FAD-monooxigenasa de <i>Mycobacterium tuberculosis</i>	*
MhpB	3-carboxietilcatecol 2,3-dioxigenasa	SGRAN_1576	MhpB (Q7N4V6, <i>Photorhabdus luminescens</i> subsp. <i>laumondii</i> , segundo resultado)	24,5
MhpC	2-hidroxi-6-oxononadienedioato hidrolasa	SGRAN_1584	MhpC (Q13QH4, <i>Paraburkholderia xenovorans</i> , quinto resultado)	32,5
HcaB	3-fenilpropionato-dihidrodol deshidrogenasa	SGRAN_1583	HcaB (B7M7P4, <i>Escherichia coli</i> O8, sexto resultado)	40,4
HcaC	3-fenilpropionato dioxigenasa (ferredoxina)	SGRAN_2802 (ThnA3)	HcaC (Q0T1X9, <i>Shigella flexneri</i> )	42,6
HcaD	3-fenilpropionato dioxigenasa (ferredoxina NAD(+) reductasa)	SGRAN_1587	HcaD (B7N6C9, <i>Escherichia coli</i> , octavo resultado)	32,3
HcaE	3-fenilpropionato dioxigenasa subunidad alfa	SGRAN_1577	HcaE (Q7N4W0, <i>Photorhabdus luminescens</i> subsp. <i>laumondii</i> )	40,8
HcaF	3-fenilpropionato dioxigenasa subunidad beta	SGRAN_1578	HcaF (Q7N4V9, <i>Photorhabdus luminescens</i> subsp. <i>laumondii</i> , octavo resultado)	34,3

\*MhpA muestra similitud con FAD-hidroxilasas (Díaz y col., 2001).

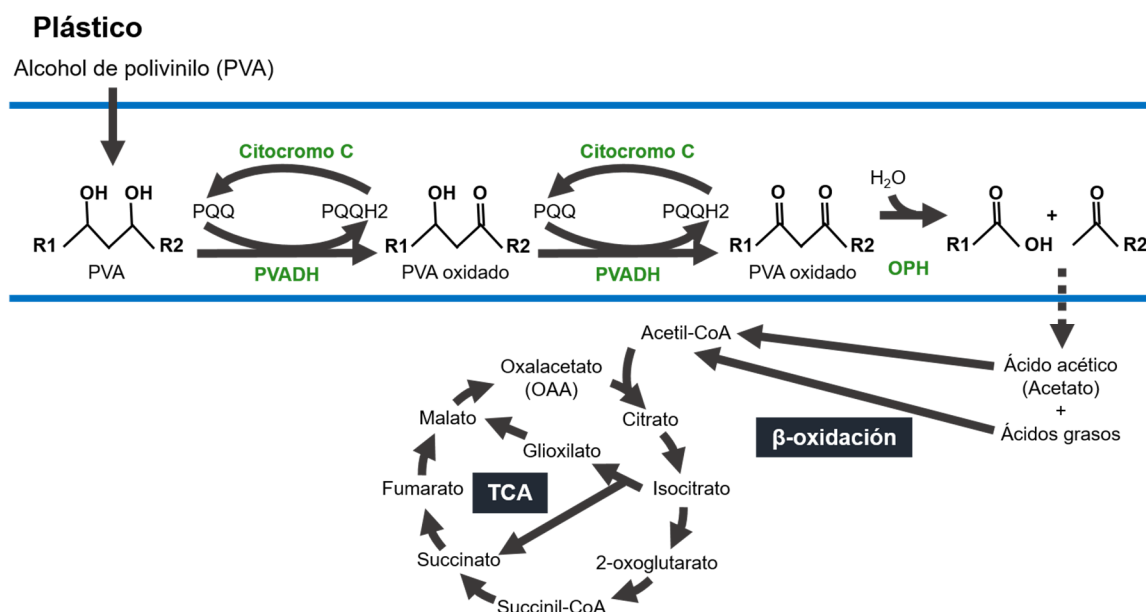


Las enzimas de la ruta baja de la degradación de 3-fenilpropanoato y 3-(3-hidroxifenil)propanoato tienen una alta identidad (>48 %) con proteínas ortólogas revisadas de la base de datos Swiss-Prot. Sin embargo, las de la ruta alta presentan una identidad más baja (del 24,5 al 42,6 %).

Continuando con las capacidades catabólicas, TFA posee los genes necesarios para la degradación del alcohol de polivinilo (PVA), como ya se mencionó previamente. La degradación de PVA en *Sphingopyxis* sp. estirpe 113P3 ocurre gracias a la intervención de tres proteínas codificadas por los genes *oph*, *pvadh* y *cytC*, los cuales se encuentran formando un operón en un mega-plásmido (Hu y col., 2008). En la estirpe 113P3, las enzimas PVA deshidrogenasa (codificada por *pvadh*) y la hidrolasa de PVA oxidado (codificada por *oph*) se encuentran en el periplasma, por lo que el polímero de PVA entra al espacio periplásmico donde es oxidado por la enzima PVA deshidrogenasa, PVADH, la cual es dependiente del cofactor pirroloquinolina quinona (PQQ) (Mamoto y col., 2006). El PVA oxidado es hidrolizado por la hidrolasa OPH (Klomklang y col., 2005) que tras varios ciclos puede generar ácido acético y ácidos grasos. El citocromo c es soluble y actúa como aceptor de electrones de PVADH en el periplasma (Mamoto y col., 2008).

Las enzimas anotadas en TFA para la degradación de PVA son muy similares a las de *Sphingopyxis* sp. estirpe 113P3. La enzima PVA deshidrogenasa de TFA, codificada por el gen SGRAN\_2681, presenta un 68 % de identidad con la enzima PVADH de *Sphingopyxis* sp. estirpe 113P3 (Uniprot: Q588Z1) y se cubre en el alineamiento el 95 % de la proteína de TFA. La hidrolasa OPH de TFA, codificada por el gen SGRAN\_2682, es un 69 % idéntica a la OPH de 113P3 (Uniprot: Q588Z2), cubriendo el 96 % de la proteína de TFA. Finalmente, el citocromo c, CytC, de TFA, codificado por el gen SGRAN\_2680, presenta un 50 % de identidad con CytC de 113P3 (Uniprot: Q588Z0), con una cobertura del 88 % de la secuencia de TFA.

A pesar de que TFA, según lo descrito anteriormente, podría degradar PVA, en la reconstrucción metabólica se determinó que esta bacteria no posee la ruta de síntesis de PQQ, por lo que las enzimas que requieran ese cofactor quedarían inutilizadas. Como consecuencia de la falta de producción de PQQ en TFA, el modelo *i*IG738 no predijo crecimiento de TFA utilizando PVA como única fuente de carbono, cuya posible vía de degradación, tras un supuesto suministro de PQQ, se ilustra en la Figura 35.



**Figura 35. Representación esquemática de la posible ruta de degradación del alcohol de polivinilo (PVA) en TFA, considerando un suministro de PQQ.** La flecha discontinua indica que finalmente, tras varios ciclos de despolimerización, se obtienen ácidos grasos y ácido acético.

### 3.2.2. Validación cuantitativa de la utilización de fuentes de carbono

Además de una predicción cualitativa del uso de las fuentes de carbono, el modelo *iIG778* puede calcular la velocidad de crecimiento de TFA en medio mínimo con una determinada fuente de carbono. Para ello, es necesario restringir el flujo de entrada de la fuente de carbono al valor de consumo experimental, medido en mmol/g (peso seco)/h. Dicho consumo se midió experimentalmente (ver materiales y métodos) monitorizando la disminución en el medio de cultivo de las siguientes fuentes de carbono: β-hidroxibutirato (β-HB), ácido sebácico, L-prolina y L-lactato. Además, se midió el peso seco de las células de TFA, en g/L, crecidas en dichas condiciones y se calculó la velocidad de crecimiento, basada en peso seco, en los distintos medios. Con los parámetros anteriores, se representó la concentración de cada fuente de carbono, en mmol/L, frente a los g/L de peso seco y la pendiente de la recta se multiplicó por la velocidad de crecimiento ( $\text{h}^{-1}$ ), obteniendo así el consumo en las unidades deseadas para incorporarlo como flujo de entrada del metabolito en el modelo.

La velocidad de crecimiento y el consumo de las fuentes de carbono se indican en la Tabla 19. Cabe destacar que, con las concentraciones utilizadas para cada una de ellas, el número de carbonos es el mismo en cada medio, de forma que, el ácido sebácico parece ser la fuente más favorable puesto que es con la que TFA alcanza una mayor velocidad de crecimiento *in vivo*.

Tabla 19. Parámetros para el cálculo del consumo de las distintas fuentes de carbono en TFA obtenidos a partir de dos réplicas biológicas.

Fuente de carbono	Pendiente recta mmol/L frente a peso seco (g/L)	Pendiente recta $\log_{10}(\text{g/L peso seco})$ frente a tiempo (min)	Velocidad de crecimiento <i>in vivo</i> ( $\text{h}^{-1}$ )	Consumo ( $\text{mmol}/(\text{g peso seco})/\text{h}$ )
Ác. Sebácico 16 mM	$-6,65 \pm 0,02$	$0,0018 \pm 0,0001$	$0,24 \pm 0,01$	$1,61 \pm 0,06$
$\beta$ -HB 40 mM	$-66,94 \pm 6,20$	$0,0014 \pm 0,0000$	$0,19 \pm 0,00$	$12,95 \pm 1,20$
L-prolina 32 mM	$-35,27 \pm 5,03$	$0,0011 \pm 0,0001$	$0,15 \pm 0,01$	$5,14 \pm 1,07$
L-lactato 53 mM	$-40,88 \pm 4,11$	$0,0005 \pm 0,0000$	$0,07 \pm 0,00$	$2,82 \pm 0,28$

Se simuló el crecimiento de TFA a través del modelo *iIG738* en las fuentes de carbono anteriores. Para ello, se restringió el flujo de entrada del metabolito al valor obtenido experimentalmente (con signo negativo) en la reacción de intercambio correspondiente EX\_bhb(e), EX\_sba(e), EX\_pro\_L(e) y EX\_lac\_L(e), para  $\beta$ -hidroxibutirato ( $\beta$ -HB), ácido sebácico, L-prolina y L-lactato, respectivamente. La gráfica de la Figura 36 muestra la comparación entre la velocidad de crecimiento calculada experimentalmente y la predicha por el modelo. Se comprueba que el mejor ajuste se obtiene para el crecimiento en ácido sebácico.

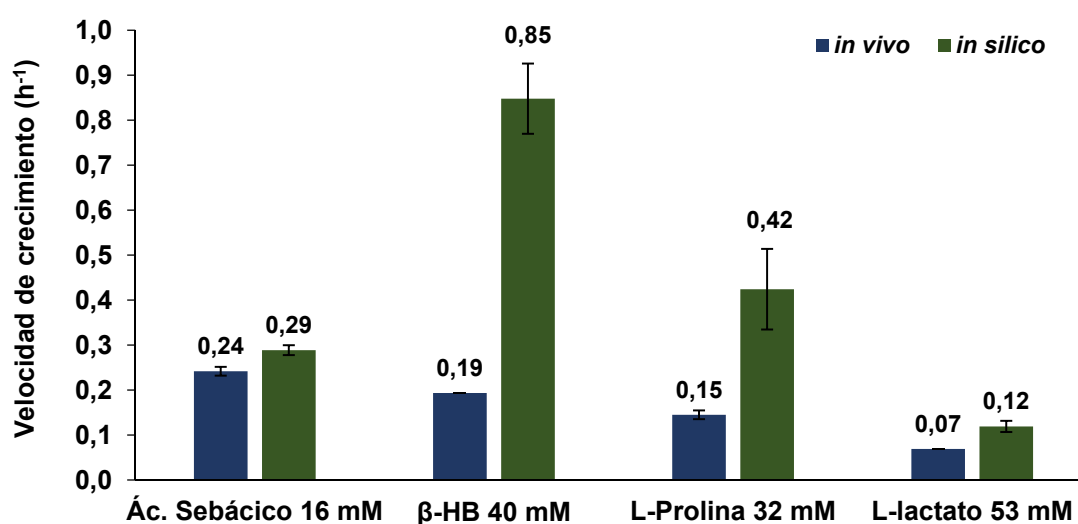


Figura 36. Velocidad de crecimiento de TFA en diferentes fuentes de carbono. En azul se muestra la calculada *in vivo* y en naranja la predicha por el modelo *iIG738* tras la introducción de la velocidad de consumo de cada fuente de carbono.

### 3.2.3. Validación del proceso de respiración anaerobia

TFA es capaz de crecer anaeróbicamente con nitrato como último aceptor de electrones (García-Romero y col., 2016). Para comprobar si el modelo *i*IG738 podía simular el crecimiento de TFA en estas condiciones, se definió un medio mínimo sin oxígeno, es decir, al límite bajo (lb) de la reacción de intercambio de oxígeno, EX\_o2(e), se le asignó 0 mmol/(g peso seco)/h y para el resto de componentes del medio mínimo, salvo el nitrato ( $\text{NO}_3$ ) y la fuente de carbono, se estableció un lb de -30 mmol/(g peso seco)/h. La concentración inicial de la fuente de carbono,  $\beta$ -HB, se fijó a 40 mM y la del último aceptor de electrones,  $\text{NO}_3$ , a 20 mM. El consumo de  $\beta$ -HB se fijó a 3 mmol/(g peso seco)/h, mientras que el del nitrato a 2,82 mmol/(g peso seco)/h, el cual fue calculado experimentalmente (Yolanda E. González, Tesis doctoral en realización). Finalmente la biomasa inicial se estableció en 0,038 g/L, según datos experimentales. Con los parámetros anteriores se simuló la acumulación de biomasa a lo largo del tiempo así como la variación de la concentración en el medio de la fuente de carbono, el  $\text{NO}_3$  y el nitrito ( $\text{NO}_2$ ).

Como se observa en la gráfica de la Figura 37, el modelo *i*IG738 puede simular la respiración anaerobia de TFA, así como la liberación al medio del  $\text{NO}_2$  producido, en proporción 1:1 con el  $\text{NO}_3$  consumido tal y como ocurre *in vivo*. Además, se establece como limitante el  $\text{NO}_3$ , puesto que el consumo de  $\beta$ -HB, disminuido por el modelo a 0,59 mmol/(g peso seco)/h, hace que la concentración de la fuente de carbono solo se haya reducido en 5 mM en el momento en el que se ha consumido todo el nitrato.

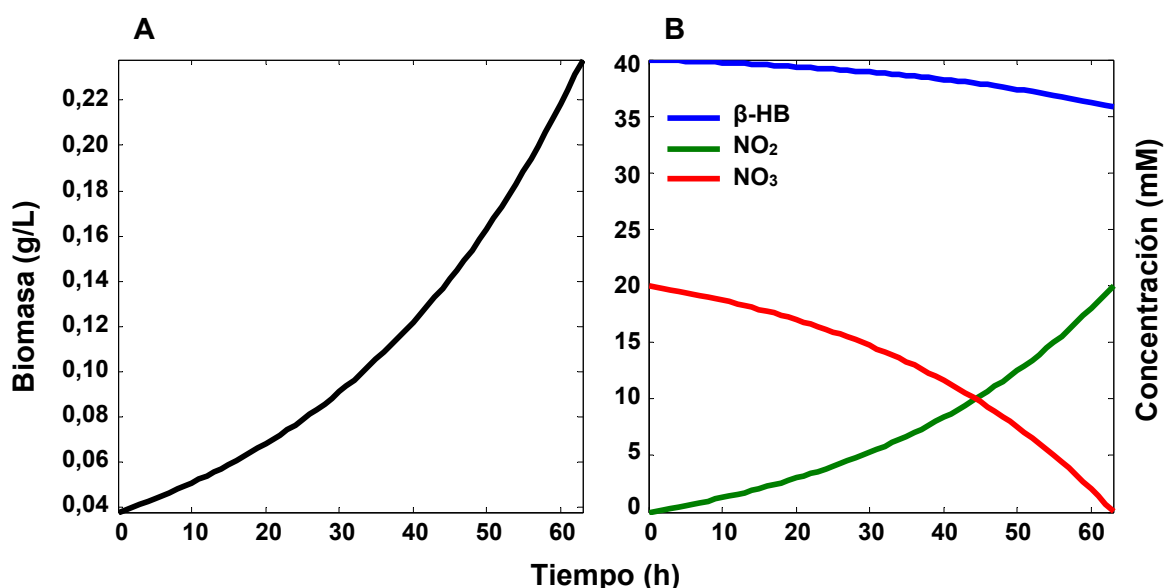
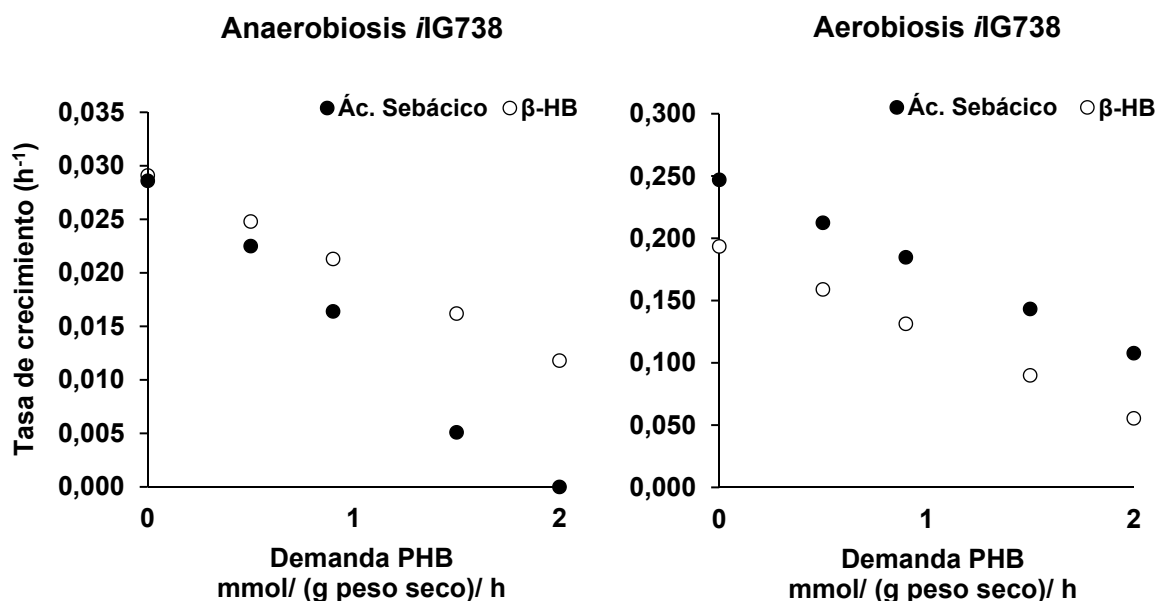


Figura 37. Optimización de la reacción de biomasa en *i*IG738, utilizando 40 mM de  $\beta$ -HB como fuente de carbono y 20 mM de nitrato ( $\text{NO}_3$ ) como último aceptor de electrones. (A) Variación de la biomasa en función del tiempo en *i*IG738, (B) Variación de la concentración de los metabolitos en el espacio extracelular en función del tiempo.

En cuanto a la tasa de crecimiento en anaerobiosis, el modelo, con las constricciones establecidas anteriormente, predijo una velocidad de crecimiento de  $0,0291 \text{ h}^{-1}$ , menor a la tasa de  $0,0553 \text{ h}^{-1}$  calculada *in vivo*.

Por otra parte, se simuló el efecto de la acumulación de gránulos de PHB sobre el crecimiento en aerobiosis y anaerobiosis utilizando ácido sebácico o  $\beta$ -HB como fuentes de carbono (Figura 38). En primer lugar, el modelo predice que, sin síntesis de gránulos de PHB, el crecimiento en aerobiosis con ácido sebácico es más rápido que con  $\beta$ -hidroxibutirato, mientras que en anaerobiosis se obtiene la misma velocidad de crecimiento. A su vez, si se incrementa la síntesis de gránulos de PHB (demanda de PHB), el modelo predice un menor crecimiento en todos los casos, lo cual es de esperar debido a que el desvío de carbono para la producción de gránulos de PHB supone una menor disponibilidad de carbono para la formación de biomasa. Por último, también se aprecia que en aerobiosis, el incremento en la demanda de gránulos de PHB tiene un efecto similar sobre el crecimiento con cualquiera de las dos fuentes de carbono, mientras que en anaerobiosis, el crecimiento con ácido sebácico es más sensible a dicha demanda.



**Figura 38. Predicción del crecimiento de TFA por el modelo iL738.** Se modela el crecimiento en ácido sebácico (flujo de consumo de  $-1,608 \text{ mmol}/(\text{g peso seco})/\text{h}$ ) y  $\beta$ -hidroxibutirato (flujo de consumo de  $-3 \text{ mmol}/(\text{g peso seco})/\text{h}$ ) tanto en aerobiosis como en anaerobiosis (flujo de  $\text{NO}_3$  de  $-2,82 \text{ mmol}/(\text{g peso seco})/\text{h}$ ) y aumentando la demanda de PHB ( $\text{DM\_PHB}$ ) de 0 a  $2 \text{ mmol}/(\text{g peso seco})/\text{h}$ . En aerobiosis el flujo de  $\text{O}_2$  se estableció a  $-30 \text{ mmol}/(\text{g peso seco})/\text{h}$  como el resto de metabolitos del medio mínimo.

## Resultados

Para confirmar el comportamiento predicho por el modelo *ilG738*, se midió la tasa de crecimiento de TFA silvestre y de la estirpe MPO209 (mutante con la inserción de un transposón en *phaC*, gen codificante de la polimerasa de PHB), tanto en anaerobiosis como aerobiosis y utilizando o ácido sebácico 16 mM o  $\beta$ -HB 40 mM como fuentes de carbono (Tabla 20).

Tabla 20. Tasa de crecimiento ( $\text{h}^{-1}$ ) de TFA silvestre y de la estirpe MPO209 (mutante en *phaC*).

Fuente de carbono	Estirpe	Anaerobiosis (20 mM $\text{NO}_3$ )	Aerobiosis
$\beta$ -HB 40 mM	TFA	$0,090 \pm 0,010$	$0,207 \pm 0,000$
	MPO209	$0,045 \pm 0,000$	$0,189 \pm 0,008$
Á. sebácico 16 mM	TFA	$0,055 \pm 0,000$	$0,235 \pm 0,000$
	MPO209	$0,042 \pm 0,000$	$0,207 \pm 0,000$

En aerobiosis, la estirpe silvestre (equivalente a las simulaciones del modelo en las que se demanda síntesis de PHB) y el mutante MPO209, que no produce PHB, tienen un comportamiento similar independientemente de la fuente de carbono que utilicen, es decir, una velocidad de crecimiento algo superior con sebácico que con  $\beta$ -HB. En condiciones de anaerobiosis, la velocidad de crecimiento del mutante de inserción en *phaC* es la misma en ambas fuentes de carbono, mientras que en la estirpe silvestre es mayor en  $\beta$ -HB que en ácido sebácico, al contrario de lo que ocurre en aerobiosis. Estos datos coinciden con la predicción del modelo *ilG738*, es decir, la velocidad de crecimiento en anaerobiosis, cuando hay síntesis de PHB, se ve más afectada al utilizar ácido sebácico como fuente de carbono que al utilizar  $\beta$ -HB, mientras que en aerobiosis, el ácido sebácico es mejor fuente de carbono haya o no producción de PHB.

#### 4. Identificación de ARNs no codificantes en *S. granuli* TFA

Además de los ARNs funcionales ribosómicos y transferentes implicados en el proceso de traducción, las células expresan otros ARNs funcionales que no se traducen y que ejercen su función como tal. Durante el proceso de anotación del genoma, se utilizó la herramienta Infernal 1.1 (Nawrocki y Eddy, 2013) y la base de datos Rfam 12.0 (Nawrocki y col., 2015) para la detección general de ARNs no codificantes. Con esta herramienta no sólo se identificaron los ARNs transferentes y ribosómicos sino también otros ARNs funcionales de tipo *riboswitch*, *housekeeping* y pequeños ARNs (Tabla 21).

Tabla 21. ARNs no codificantes anotados en el genoma de TFA.

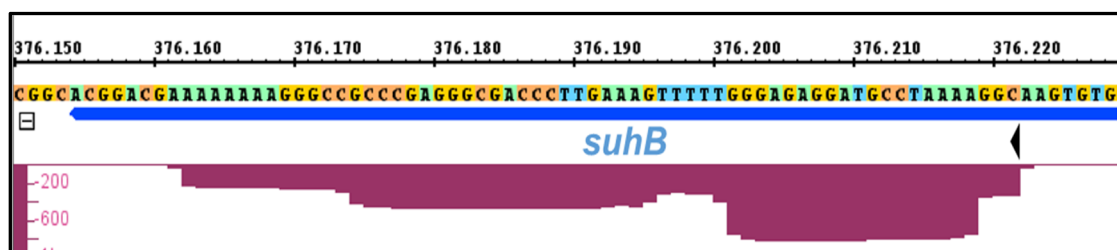
ARN no codificante	Tipo	Función
<b>SGRAN_0363 (<i>suhB</i>)</b>	Pequeño ARN	Desconocida
<b>SGRAN_1306 (<i>serC</i>)</b>	Pequeño ARN	Bloquea la traducción del gen <i>serC</i> (Corbino y col., 2005)
<b>SGRAN_1326 (<i>ptaRNA1</i>)</b>	Pequeño ARN	ARN regulador antisentido (Findeiss y col., 2010)
<b>SGRAN_1330 (<i>traJ-II</i>)</b>	Pequeño ARN	Podría tratarse de un ARN antisentido del gen <i>traJ</i> (Weinberg y col., 2010)
<b>SGRAN_1524</b>	<i>Riboswitch</i> de glicina	Implicado en la degradación de glicina (Mandal y col., 2004)
<b>SGRAN_2154 (6S/<i>ssrS</i>)</b>	<i>Housekeeping</i>	Reprime la expresión de promotores dependientes del factor $\sigma^{70}$ durante la fase estacionaria (Wassarman y Storz, 2000)
<b>SGRAN_2431 (<i>srp</i>)</b>	<i>Housekeeping</i>	Partícula de reconocimiento de señal. Forma parte de un complejo ribonucleico que se une al ribosoma durante la síntesis y translocación de proteínas secretoras (Zwieb y col., 2005)
<b>SGRAN_2950 (<i>rnpB</i>)</b>	<i>Housekeeping</i>	Codifica para un ARN catalítico que es capaz de cortar otros ARN. Forma parte de la ribonucleasa P (Kole y Altman, 1979)
<b>SGRAN_3138 (<i>tmRNA</i>)</b>	<i>Housekeeping</i>	Molécula híbrida que tiene propiedades de ARNm y ARNt, interviene en la <i>trans</i> -traducción (Macé y Gillet, 2016)
<b>SGRAN_3624</b>	<i>Riboswitch</i> de cobalamina	Implicado en la regulación del metabolismo de la vitamina B12 (Vitreschak y col., 2003)
<b>SGRAN_3917</b>	<i>Riboswitch</i> TPP	Controla la expresión de genes implicados en la síntesis o transporte de tiamina y sus derivados fosforilados (Haller y col., 2013)

La función de todos los ARNs no codificantes detectados pudo identificarse gracias a lo descrito en las bases de datos y la bibliografía, salvo la del pequeño ARN *SuhB*.

## 5. Caracterización del pequeño ARN regulador SuhB

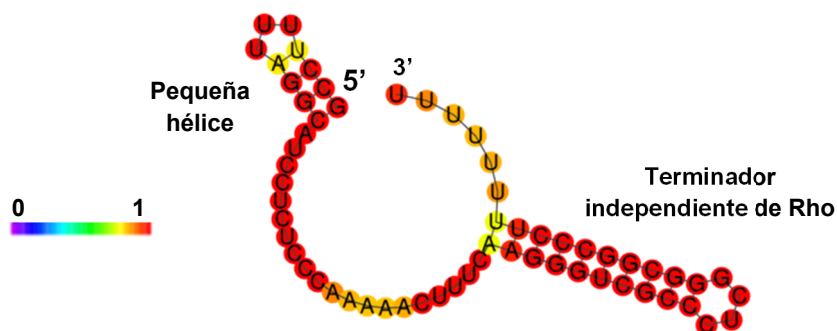
### 5.1. Estudio de la conservación de la secuencia y el contexto génico de SuhB en el género *Sphingopyxis*

El pequeño ARN SuhB, clasificado en la familia RF00519 de la base de datos Rfam, se identificó en *Sphingopyxis granuli* estirpe TFA tanto *in silico*, utilizando la herramienta Infernal, como experimentalmente mediante experimentos de transcriptómica y la determinación del inicio de la transcripción (TSS) (Figura 39).



**Figura 39. Detección *in silico* y experimental del pequeño ARN SuhB en TFA.** Se muestra la secuencia del genoma de TFA en cuya cadena complementaria se encuentra codificado SuhB. La predicción de Infernal se representa en azul y, en morado, la expresión del pequeño ARN medida por la acumulación de lecturas de la secuenciación de una genoteca de ADNc obtenida a partir de células en fase exponencial de crecimiento en medio con  $\beta$ -HB 40 mM. La flecha negra marca el TSS determinado experimentalmente (ver apartado 5.2 de resultados).

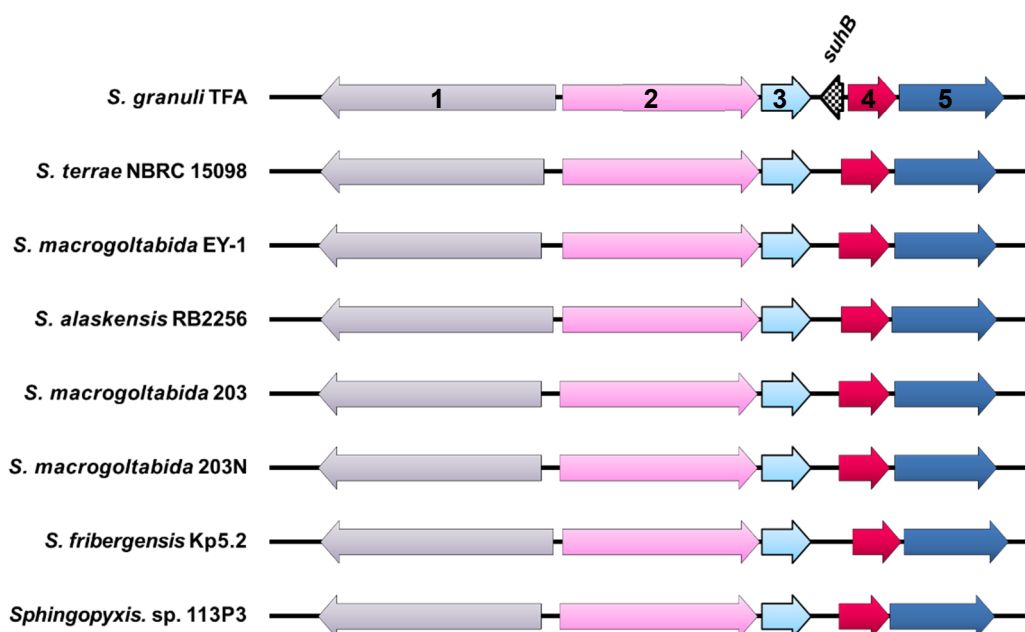
En la Figura 39 se aprecia que el transcrito obtenido experimentalmente comienza en la octava base del gen *suhB* anotado por Infernal. Además, el fin del transcrito coincide con el final de la ristra de Us, como suele ocurrir con los terminadores independientes del factor Rho, mientras que la predicción de Infernal contiene siete nucleótidos más. Como consecuencia, se estableció como correcta la secuencia definida por el experimento de transcriptómica para el gen *suhB*, de 61 nucleótidos, en lugar de la predicha por Infernal, de 75 nucleótidos. Este pequeño ARN contiene dos estructuras bien diferenciadas: una pequeña hélice en su extremo 5' y una secuencia terminadora independiente de Rho hacia su extremo 3' (Figura 40).



**Figura 40. Estructura de la secuencia del pequeño ARN SuhB.** Los colores indican la probabilidad de apareamiento de las bases, según la escala de 0 a 1 representada en la zona inferior de la Figura (en el caso de las bases libres se representa la probabilidad de que se encuentren sin hibridar).

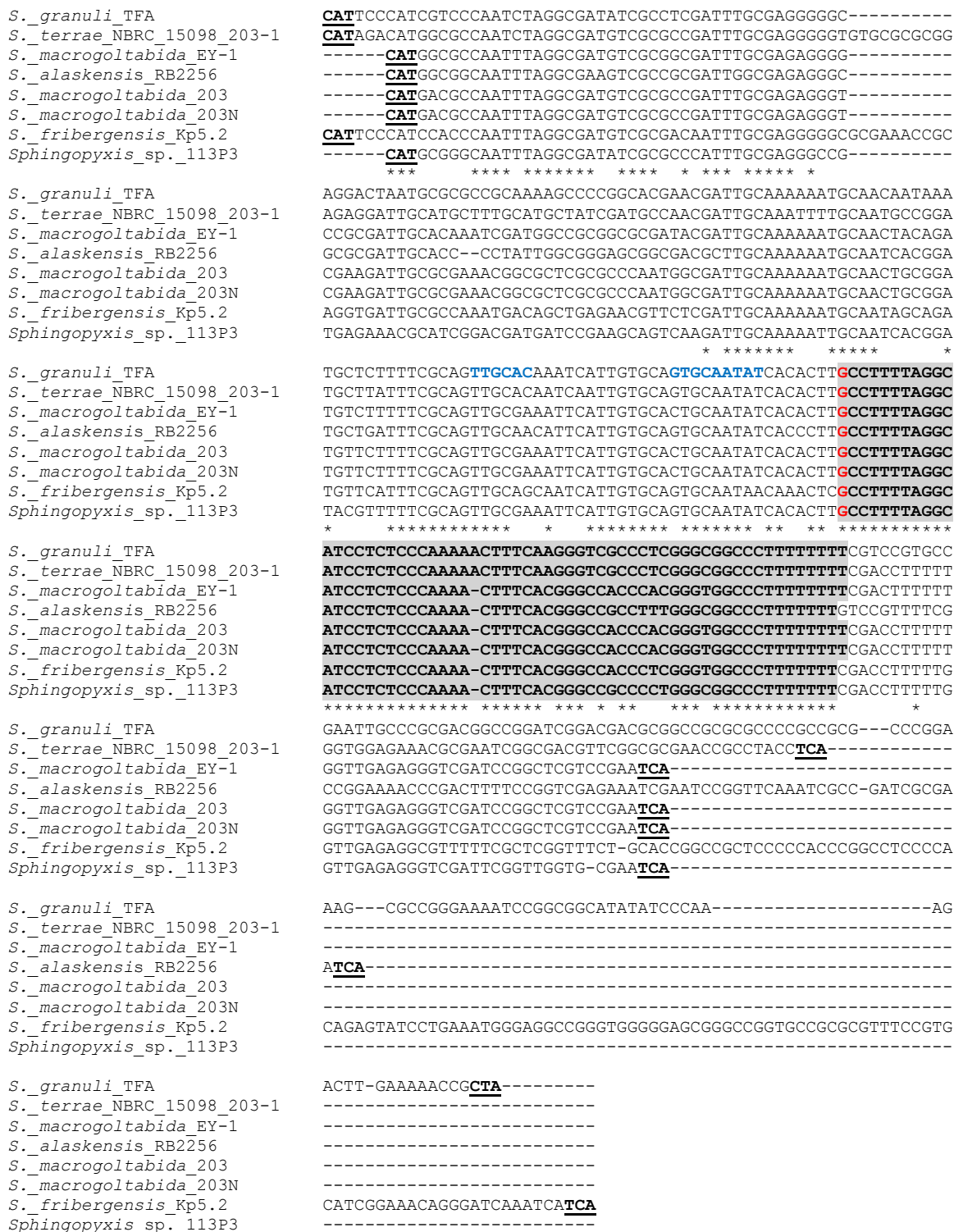


Según la información disponible en Rfam, SuhB se encuentra altamente conservado dentro de las  $\alpha$ -proteobacterias, incluso en múltiples copias en algunos genomas como los de *Bradyrhizobium japonicum* (5 copias) y *Caulobacter crescentus* (4 copias) (Corbino y col., 2005). Sin embargo, dentro del género *Sphingopyxis*, TFA es el único genoma donde se ha anotado el gen *suhB*. Para detectar la presencia de *suhB* en bacterias del género *Sphingopyxis* con genoma secuenciado y cerrado en el momento del estudio, se utilizó el servidor web SyntTax (Oberto, 2013) que posibilita la identificación de zonas con la misma organización génica en diferentes genomas. Utilizando la secuencia de la proteína codificada por el gen SGRAN\_0362 (gen número 3 de la Figura 41) como dato de entrada para SyntTax, se evaluó la conservación de dicha proteína y de las codificadas por los genes adyacentes, desde SGRAN\_0360 a SGRAN\_0365. El programa detectó sintenia de todos los genes y en todos los genomas estudiados (Figura 41). Además, se detectó una región intergénica entre los genes 3 y 4, donde se encuentra el gen *suhB* en TFA, de tamaño suficiente para albergar un gen que codificara el pequeño ARN en las otras *Sphingopyxis*.



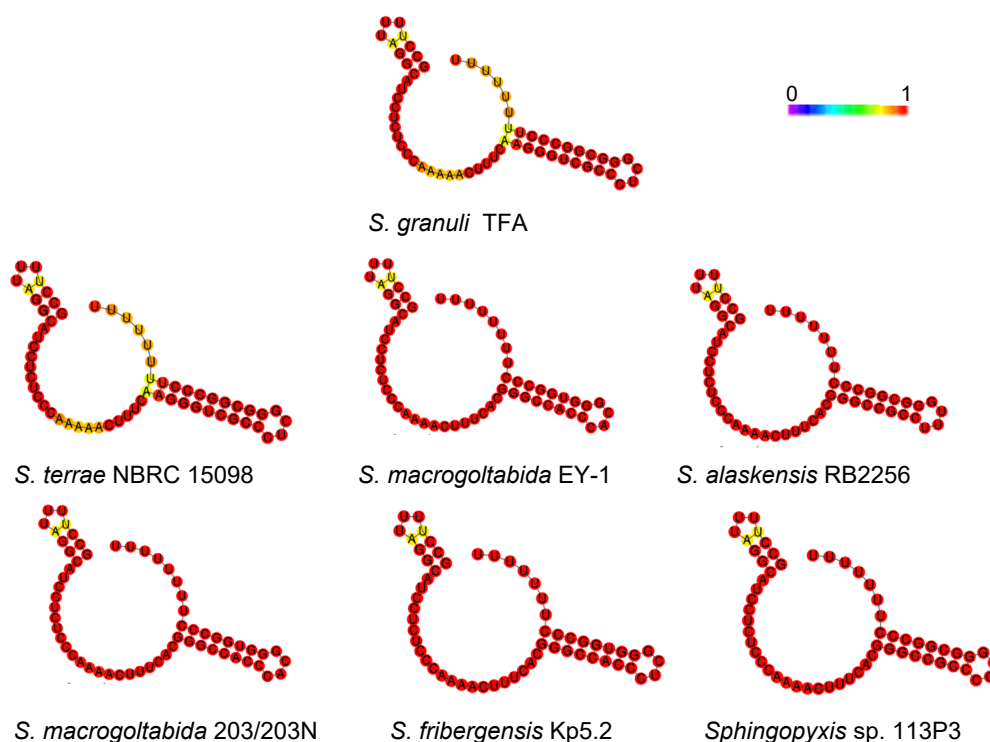
**Figura 41. Conservación del contexto génico de *suhB* en los genomas de bacterias del género *Sphingopyxis*.** En todas las especies, se representan genes que codifican: (1) oligopeptidasa B (SGRAN\_0360), (2) X-Pro aminopeptidasa (SGRAN\_0361), (3) 4-hidroxibenzoil-CoA tioesterasa (SGRAN\_0362), (4) proteína no caracterizada (SGRAN\_0364) y (5) riboflavina desaminasa (*ribD*, (SGRAN\_0365). La Figura está adaptada del resultado obtenido con la herramienta SyntTax.

La comparación de las secuencias de la región intergénica, desde el codón de fin del gen 3 al codón de inicio del gen 4, permitió detectar una zona altamente conservada y muy similar al gen *suhB* de TFA, no solo en la región transcrita sino en la posible región promotora (Figura 42). Además, todas estas secuencias se localizaban en orientación divergente al gen 4.



**Figura 42. Alineamiento de la región intergénica que contiene a *suhB* en TFA y la región homóloga en el resto de *Sphingopyxis*.** Se marca en gris la secuencia de *suhB* definida por transcriptómica en TFA, en la cadena en la que está codificado el pequeño ARN, y la región que alinea con ella en las otras bacterias. En rojo se indica la posición en la que se predice el inicio de la transcripción de SuhB en TFA, que se encuentra conservada en todas las bacterias del género *Sphingopyxis*. La posible secuencia promotora, definida a través del programa BPROM (Solovyev y Salamov, 2011), se marca en azul. Subrayado y en negrita se muestran los codones de inicio y stop de los genes adyacentes. Los asteriscos denotan conservación de nucleótidos.

Utilizando la herramienta web RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) (Lorenz y col., 2011) se comprobó que la conservación en secuencia se traducía en una conservación de la estructura del posible pequeño ARN SuhB en las distintas *Sphingopyxis* analizadas. En todas ellas, se conserva la pequeña hélice en el extremo 5' y el terminador intrínseco en el 3' (Figura 43).

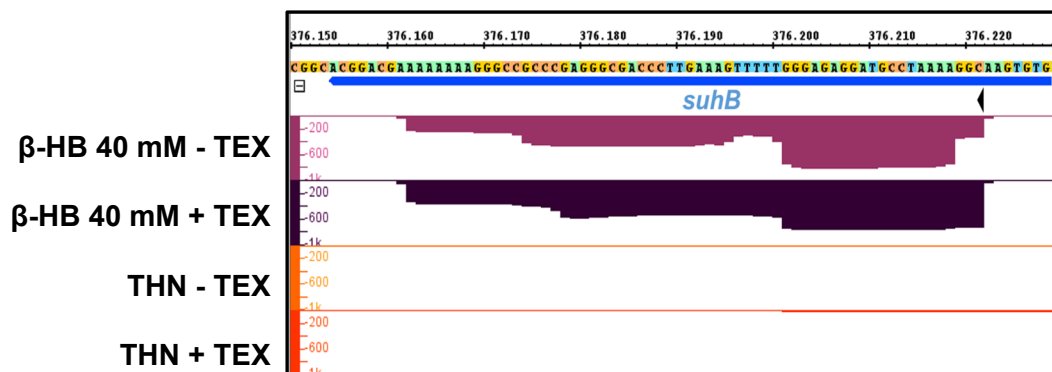


**Figura 43. Conservación de la estructura de SuhB en las distintas *Sphingopyxis*.** El color, según la escala indicada, representa la probabilidad de emparejamiento de las bases o, en su caso, la probabilidad de que permanezcan sin aparear.

## 5.2. Expresión diferencial de *suhB*

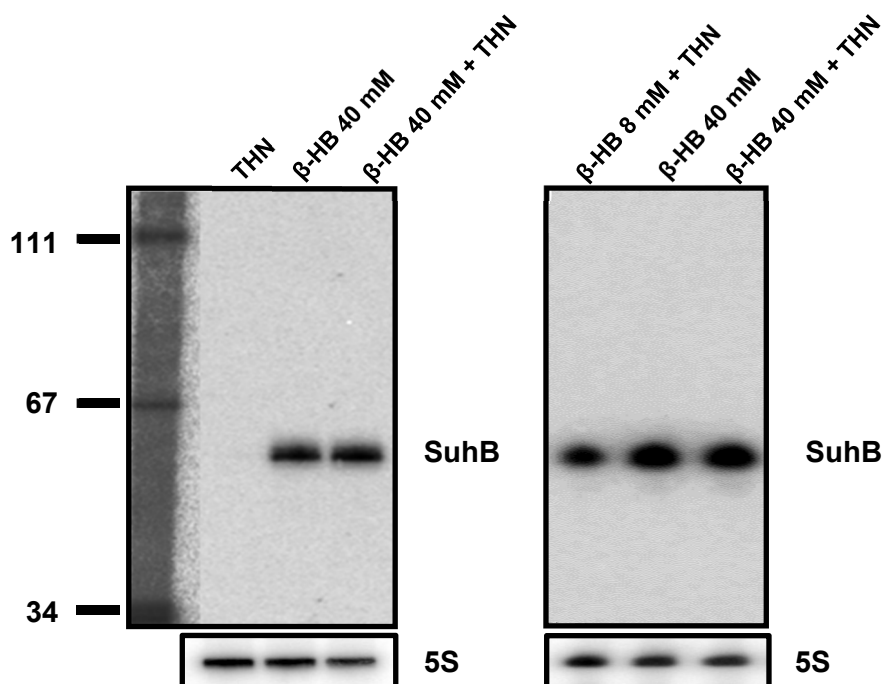
Con el fin de identificar genes con expresión diferencial en células de TFA cuando crece en una fuente de carbono preferencial, como el  $\beta$ -HB, o en el solvente orgánico tetralina, y detectar los inicios de transcripción (TSS), se realizaron experimentos de dRNA-seq (ver materiales y métodos). Para evaluar la expresión diferencial entre las dos fuentes de carbono, se analizaron las muestras sin tratar con TEX. Se encontró que el gen *suhB* sólo se expresa en células creciendo exponencialmente en  $\beta$ -HB, como puede apreciarse en la Figura 44 extraída de la visualización de los datos de expresión obtenidos con el software IGB (Integrated Genome Browser) (Freese y col., 2016).

## Resultados



**Figura 44. Expresión diferencial de *suhB* en  $\beta$ -HB y tetralina.** En la representación, el gen *suhB* predicho por Infernal 1.1 se encuentra codificado en la cadena negativa (línea azul). Con una flecha se señala el inicio de la transcripción predicho por TSSpredator (Dugar y col., 2013). En diferente color se muestra, normalizada para poder comparar las diferentes muestras, la acumulación de lecturas tras la secuenciación de las genotecas de ADNc obtenidas en cada condición.

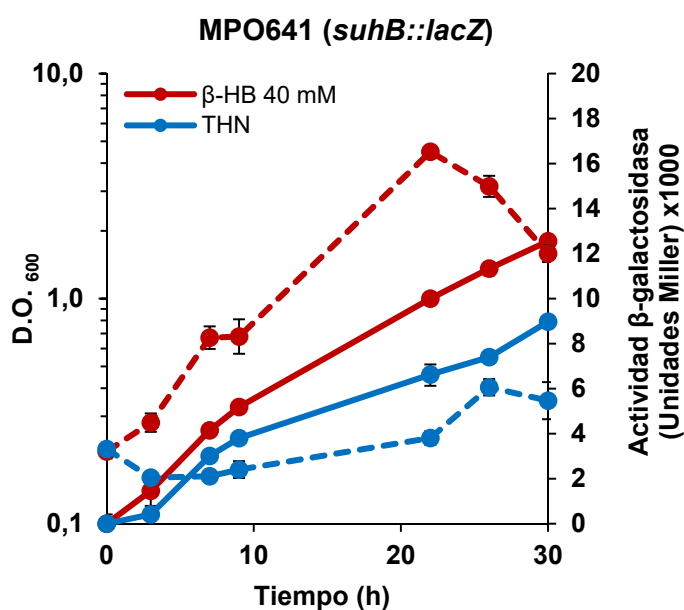
Por otra parte, con los datos obtenidos en el experimento de dRNA-seq y mediante el programa TSSpredator, se detectó el inicio de la transcripción de *suhB*. Para confirmar el tamaño de SuhB, se detectó el sRNA mediante *Northern Blot* empleando como sonda el oligonucleótido *suhB\_northern*, que es complementario al extremo 5' del pequeño ARN. Como se observa en la Figura 45, SuhB solo se expresa si en el medio está presente el  $\beta$ -HB, independientemente de que haya o no tetralina.



**Figura 45. Detección de SuhB mediante *Northern Blots*.** Se muestra el resultado de la hibridación de la sonda radiactiva complementaria a SuhB en muestras de ARN extraídas de células de TFA creciendo en distintos medios en fase exponencial. El marcador de tamaño (en nucleótidos) se muestra a la izquierda.

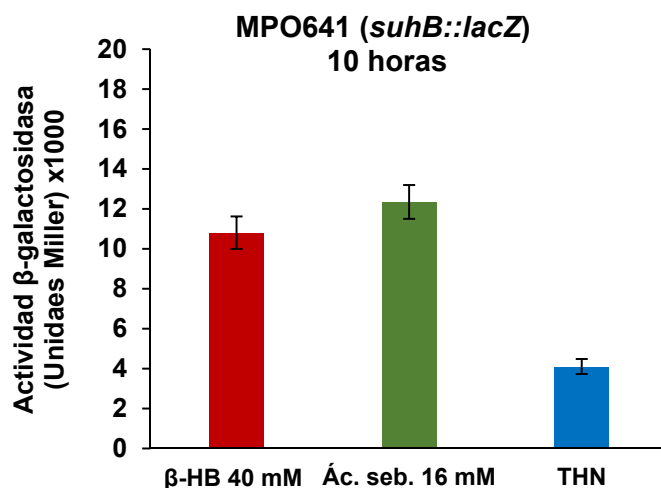
Además, la señal correspondiente al transcrito de *suhB* se detectó ligeramente por debajo de la banda de 67 nucleótidos del marcador de tamaño, lo cual se ajusta más al tamaño del ARN detectado mediante dRNA-Seq que a la anotación realizada por Infernal.

Para ampliar el estudio de la expresión diferencial de *suhB* de manera más sencilla, se construyó una fusión transcripcional *PsuhB::lacZ* en el vector pIC552 (el fragmento clonado en dicho vector contiene el posible promotor de *suhB* y finaliza dos nucleótidos aguas arriba del TSS de *suhB*). Se obtuvo de esta forma el plásmido pMPO1151, el cual se integró en el cromosoma de TFA generándose la estirpe MPO641. La expresión del promotor de *suhB* se cuantificó midiendo la actividad  $\beta$ -galactosidasa en células de TFA creciendo en aerobiosis en medio mínimo con  $\beta$ -HB 40 mM o con tetralina como única fuente de carbono (Figura 46). Los cultivos se inocularon a partir de células adaptadas al crecimiento en tetralina como se describe en la Figura 17 del apartado 8.1.1 de materiales y métodos. Según los datos obtenidos, la expresión de *suhB* aumenta en células creciendo con  $\beta$ -HB como fuente de carbono.



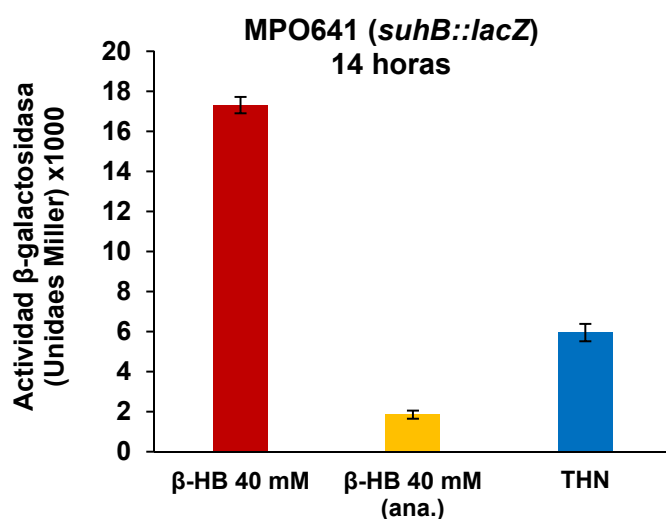
**Figura 46. Expresión de *suhB* en células de TFA crecidas en  $\beta$ -HB 40 mM o THN.** Los cultivos se iniciaron partiendo de células crecidas en medio mínimo con THN hasta fase exponencial. Se representa la cinética de expresión de *PsuhB::lacZ* (actividad  $\beta$ -galactosidasa en línea discontinua) a lo largo de la curva de crecimiento (medido por incremento de densidad óptica, en línea continua).

Con objeto de averiguar si el  $\beta$ -HB era un inductor específico o no, se midió la actividad  $\beta$ -galactosidasa en células de MPO641 crecidas durante 10 horas en aerobiosis en medio mínimo con ácido sebácico 16 mM como fuente de carbono (Figura 47). De nuevo los cultivos se inocularon a partir de células adaptadas al crecimiento en tetralina.



**Figura 47. Expresión de *suhB* en células de TFA crecidas en  $\beta$ -HB 40 mM, ácido sebácico 16 mM o THN.** Los cultivos se iniciaron partiendo de células crecidas en medio mínimo con THN hasta fase exponencial. Se representa la expresión de *P<sub>suhB</sub>::lacZ* en células crecidas aeróbicamente durante 10 horas en medio mínimo con las fuentes de carbono indicadas.

Como se observa, *suhB* se induce en presencia de una buena fuente de carbono, ya sea  $\beta$ -HB 40 mM o ácido sebácico 16 mM. Debido a la menor velocidad de crecimiento de TFA utilizando tetralina, cabía la posibilidad de que la inducción de *suhB* se debiera al estado energético de la célula. Para comprobar esta hipótesis, se midió la expresión de *P<sub>suhB</sub>::lacZ* en células de MPO641 crecidas durante 14 horas en condiciones anaeróbicas en medio mínimo con  $\beta$ -HB 40 mM con nitrato 20 mM como aceptor de electrones (Figura 48). También en este caso los cultivos se inoculaban a partir de células adaptadas al crecimiento en tetralina.

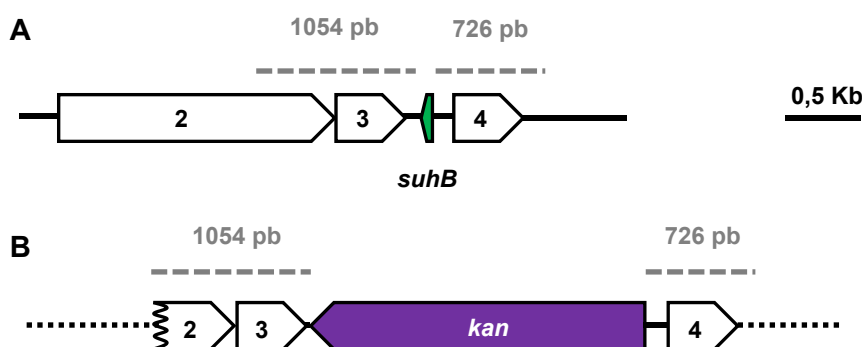


**Figura 48. Expresión de *suhB* en células de TFA.** Las células se cultivaron aeróbicamente en medio mínimo con  $\beta$ -HB 40 mM o tetralina y anaeróbicamente (ana.) en medio mínimo con  $\beta$ -HB 40 mM y nitrato 20 mM como aceptor de electrones. En todos los casos los cultivos se iniciaron partiendo de células crecidas en medio mínimo con tetralina hasta fase exponencial. Se representa la actividad  $\beta$ -galactosidasa a las 14 horas de crecimiento.

Como muestra la Figura 48, no se observa inducción de la expresión de *suhB* en condiciones anaeróbicas, lo que sugiere que la expresión del pequeño ARN estaría determinada por el rendimiento energético que la célula puede obtener en un medio de cultivo concreto en lugar de por la fuente de carbono disponible.

### 5.3. Construcción de un mutante carente del pequeño ARN SuhB

Para caracterizar la función de SuhB, se construyó un mutante de sustitución del correspondiente gen por un gen de resistencia a kanamicina y se analizó su fenotipo. Para la construcción del mutante se empleó el plásmido pMPO1150, basado en el plásmido suicida pEX18tc. Este plásmido contiene el gen *sacB*, el cual confiere sensibilidad a sacarosa a las células en las que se expresa. En pEX18tc se clonaron dos fragmentos correspondientes a las secuencias aguas arriba y aguas abajo del gen *suhB* (Figura 49A), entre las cuales se clonó la resistencia a kanamicina (Figura 49B).



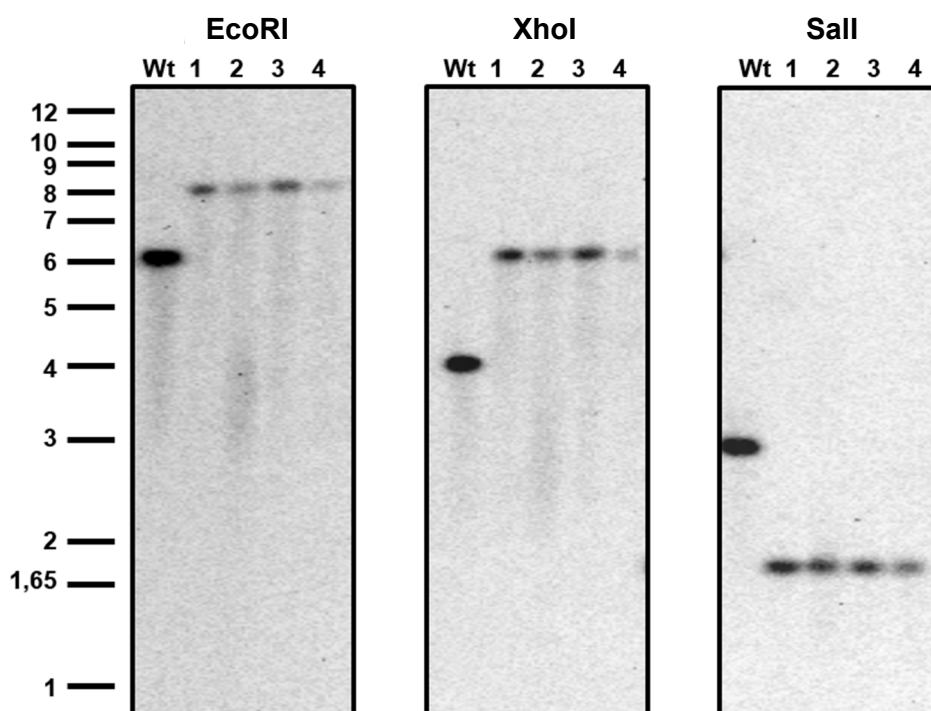
**Figura 49. Localización de *suhB* en el genoma de TFA (A) y esquema de la construcción del plásmido pMPO1150 (B).** Los números 2, 3 y 4, de forma análoga a la Figura 41 (sintenia), hacen referencia a genes codificantes para (2) X-Pro aminopeptidasa (SGRAN\_0361), (3) 4-hidroxibenzoil-CoA tioesterasa (SGRAN\_0362) y (4) una proteína no caracterizada (SGRAN\_0364). El plásmido pMPO1150 deriva del vector pEX18tc y contiene dos regiones de homología (marcadas en línea discontinua) que corresponden a las secuencias aguas arriba y aguas abajo del gen *suhB*. Entre ellas, se encuentra un gen de resistencia a kanamicina.

El plásmido pMPO1150 se introdujo en células de TFA a través de conjugación triparental y se seleccionaron los transconjugantes en medio rico con tetraciclina (resistencia del plásmido) y kanamicina (resistencia para la sustitución del gen *suhB*), con el fin de obtener candidatos con el plásmido integrado en el cromosoma por una de las dos regiones homólogas. La presencia del vector en los candidatos seleccionados se comprobó amplificando una región interna de 782 pares de bases del gen de resistencia a kanamicina, utilizando la pareja de oligonucleótidos KmFw\_pk18 y KmRev\_pk18. Al tratarse de un vector

## Resultados

no replicativo en TFA, éste debería estar integrado en el cromosoma en aquellos candidatos resistentes a tetraciclina y kanamicina.

Tras un periodo de crecimiento en medio no selectivo, los candidatos se sembraron en placas con kanamicina y sacarosa para seleccionar aquellos que, gracias al segundo evento de recombinación, hubiesen perdido el esqueleto del plásmido pMPO1150, que contiene el gen *sacB*, manteniendo el gen de resistencia a kanamicina. La correcta sustitución de *suhB* por el gen de resistencia a kanamicina se comprobó mediante *Southern Blot* (Figura 50) en 4 de los candidatos seleccionados. Para dicha comprobación, se extrajo ADN cromosómico de la estirpe silvestre de TFA y de los 4 candidatos que podrían ser mutantes de *suhB*. Posteriormente, todas las muestras se digirieron, en reacciones de restricción distintas, con EcoRI, XhoI y Sall. En la hibridación, se usó como sonda el fragmento de 726 pb aguas arriba al gen *suhB* que se amplificó para la construcción del mutante. Se esperaba una banda de 6 kb en el genoma de la estirpe silvestre y una de 8 Kb en el genoma de los candidatos positivos en la digestión con EcoRI, una banda de 3,8 Kb para la estirpe silvestre y de 5,9 Kb para los candidatos en la digestión con XhoI y, finalmente, una banda de 2,8 Kb para la estirpe silvestre y de 1,7 para los candidatos en la restricción con Sall. En la Figura 50 se muestran las señales obtenidas en el *Southern Blot* descrito anteriormente, las cuales confirmaron que los cuatro candidatos analizados eran mutantes de sustitución de *suhB*. Finalmente, se seleccionó el candidato 1 y el mutante se denominó MPO642.



**Figura 50. Southern Blot de ADN cromosómico de TFA silvestre y los candidatos  $\Delta suhB::kan$ .** TFA silvestre se indica como wt y los candidatos se numeran del 1 al 4. A la izquierda de la Figura se representa el marcador de peso molecular, en kb.



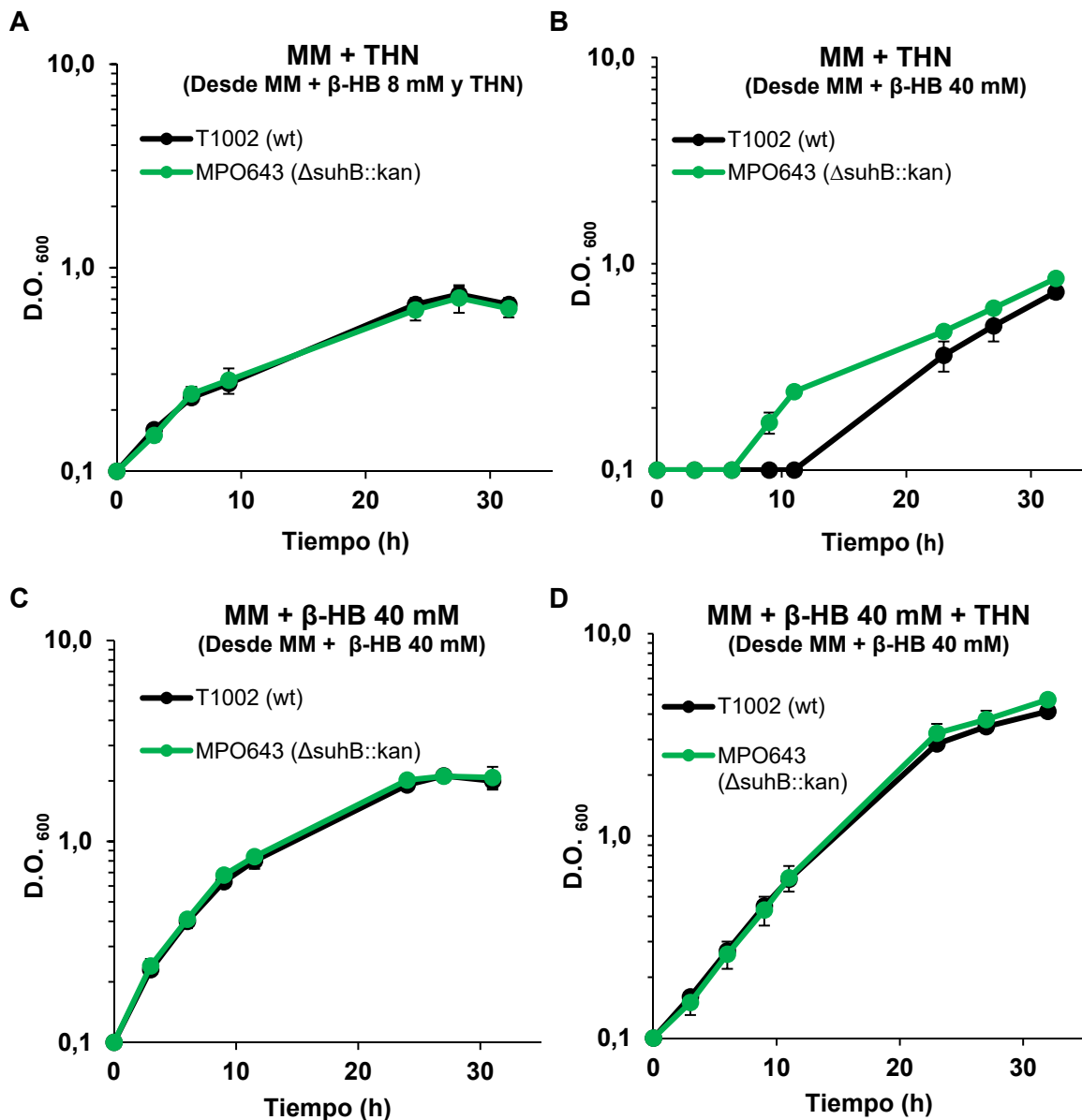
Para evaluar la expresión de los genes *thn* en el mutante de *suhB*, se insertó en el cromosoma de la estirpe MPO642 el plásmido pIZ1002, el cual porta una fusión traduccional *thnC::lacZ* (Martínez-Pérez y col., 2004). La estirpe obtenida se nombró MPO643.

### 5.3.1. Caracterización del crecimiento de la estirpe MPO643

Debido a la expresión diferencial del pequeño ARN *suhB* en presencia de  $\beta$ -HB (con o sin tetralina) observada en los experimentos de dRNA-seq, los ensayos de *Northern Blot* y las medidas de la actividad  $\beta$ -galactosidasa de la fusión *PsuhB::lacZ*, cabría esperar que un mutante carente de *suhB* tuviese afectado el crecimiento, con respecto a TFA silvestre, según se cultivase en las diferentes fuentes de carbono.

Para comprobar dicha hipótesis, se cultivaron las cepas MPO643 y T1002 (silvestre con la fusión *thnC::lacZ* integrada en el cromosoma) en medio mínimo con tetralina,  $\beta$ -HB 40 mM o en presencia de ambas fuentes de carbono simultáneamente. Los cultivos se iniciaron a partir de células en fase exponencial (a una D.O.<sub>600</sub> de ~0,8) crecidas en medio mínimo con  $\beta$ -HB 40 mM o con  $\beta$ -HB 8 mM y tetralina. En este último caso, las células del preinóculo ya estarían creciendo a expensas de la tetralina. Las gráficas representadas en la Figura 51 muestran el crecimiento de las estirpes MPO643 y T1002 en medio mínimo con tetralina (A y B),  $\beta$ -HB 40 mM (C) o ambas fuentes de carbono (D).

## Resultados



**Figura 51. Curvas de crecimiento de T1002 (TFA; *thnC::lacZ*) y MPO643 ( $\Delta$ suhB::kan; *thnC::lacZ*) en distintos medios.** (A) Crecimiento en medio mínimo con THN de cultivos que se iniciaron a partir de preinóculos en medio mínimo con  $\beta$ -HB 8 mM y THN, (B) Crecimiento en medio mínimo con THN de cultivos que se iniciaron a partir de preinóculos en medio mínimo con  $\beta$ -HB 40 mM. (C) Crecimiento en medio mínimo con  $\beta$ -HB 40 mM y (D) medio mínimo con  $\beta$ -HB 40 mM y THN de cultivos que se iniciaron a partir de células crecidas en medio mínimo con  $\beta$ -HB 40 mM.

El patrón de crecimiento de ambas estirpes es el mismo en todos los medios salvo en tetralina cuando el preinóculo ha sido crecido en medio mínimo con  $\beta$ -HB 40 mM (Figura 51B). En este caso, la estirpe MPO643 presenta una fase de latencia menor (unas 6 horas) que la estirpe silvestre T1002 (más de 10 horas).

### 5.3.2. Caracterización de la inducción de los genes *thn* en el mutante MPO643 en medio mínimo con tetralina

La ausencia del pequeño ARN SuhB afecta al crecimiento de TFA en medio mínimo con tetralina, disminuyendo la fase de latencia cuando los preinóculos no están adaptados al crecimiento en esta fuente de carbono. Este hecho podría indicar una alteración de la expresión de los genes *thn* en estas condiciones en la estirpe MPO643 con respecto a la estirpe silvestre T1002. Para corroborar dicha hipótesis, se midió la inducción de la fusión *thnC::lacZ* a lo largo del tiempo en condiciones de inducción de los genes *thn* (medio mínimo con tetralina) en las estirpes T1002 y MPO643 mediante ensayos de actividad  $\beta$ -galactosidasa. Los cultivos se inocularon a partir de células en fase exponencial cultivadas en condiciones de no inducción de los genes *thn* (medio mínimo con  $\beta$ -HB 40 mM).

Los resultados obtenidos (Figura 52) muestran que los genes *thn* se inducen antes y a mayor nivel en la estirpe mutante MPO643 que en TFA silvestre cuando crecen en medio mínimo con tetralina, lo que explicaría la reducción de la fase de latencia observada anteriormente.

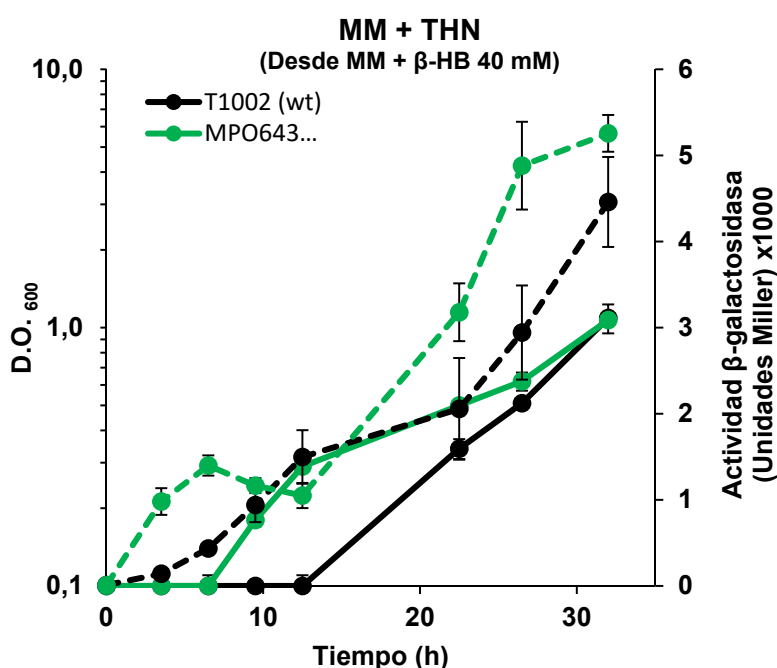


Figura 52. Curva de crecimiento y expresión de la fusión traduccional *thnC::lacZ* en MM con tetralina en las estirpes T1002 (TFA; *thnC::lacZ*) y MPO643 ( $\Delta$ *suhB::kan*; *thnC::lacZ*). El crecimiento se representa en línea continua y la actividad  $\beta$ -galactosidasa en línea discontinua.

### 5.3.3. Caracterización de la inducción de los genes *thn* en el mutante MPO643 en condiciones de represión catabólica

Como se explicó en la introducción de esta Tesis, la inducción de los genes de degradación de tetralina se inhibe cuando en el medio está presente, junto con la tetralina, una fuente de carbono preferencial. Una de estas fuentes de carbono es el  $\beta$ -HB, que, a una concentración de 40 mM, provoca la represión de los genes *thn* durante las primeras 15-20 horas de crecimiento. Debido a que la ausencia de SuhB en medio mínimo con tetralina (condiciones de inducción) hace que los genes *thn* se induzcan antes y a mayor nivel, cabe esperar que en medio mínimo con tetralina y  $\beta$ -HB 40 mM (condiciones de represión catabólica) la ausencia de SuhB tenga un efecto sobre la expresión de los genes *thn*. Para comprobarlo, se realizaron curvas de crecimiento y se midió la actividad  $\beta$ -galactosidasa de la fusión *thnC::lacZ* en las estirpes T1002 y MPO643 crecidas en medio mínimo con tetralina y  $\beta$ -HB 40 mM. Los preinóculos con los que se iniciaron las curvas provenían de cultivos en fase exponencial crecidos en medio mínimo con  $\beta$ -HB 40 mM (condiciones de no inducción de los genes *thn*).

El resultado del crecimiento y la inducción de los genes *thn* en las estirpes T1002 y MPO643 se ilustra en la Figura 53.

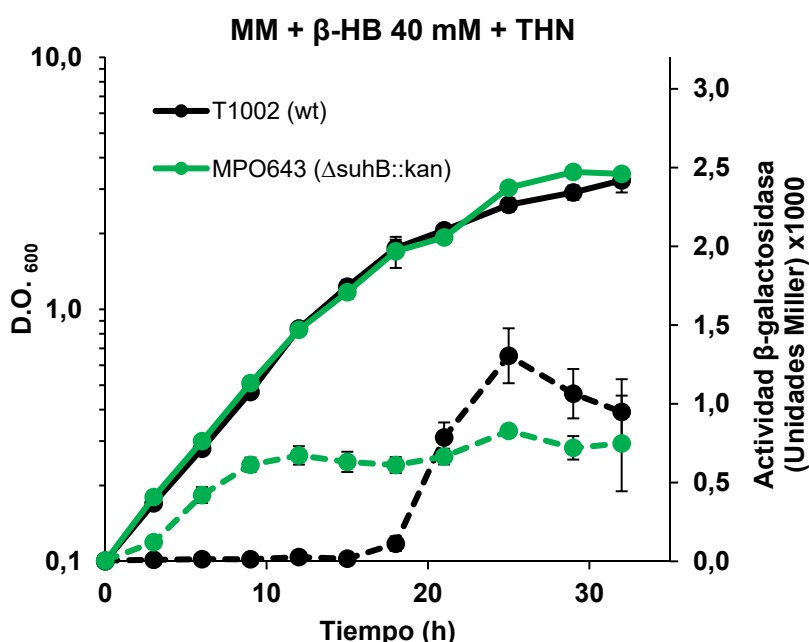
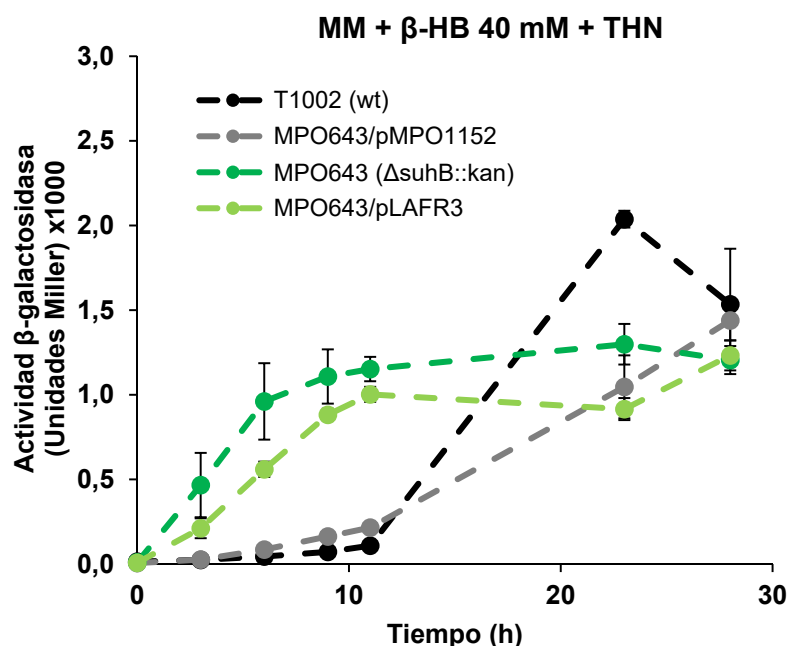


Figura 53. Curva de crecimiento y expresión de la fusión traduccional *thnC::lacZ* en MM con tetralina y  $\beta$ -HB 40 mM en las estirpes T1002 (TFA; *thnC::lacZ*) y MPO643 ( $\Delta$ suhB::kan; *thnC::lacZ*). El crecimiento se representa en línea continua y la actividad  $\beta$ -galactosidasa en línea discontinua.

Mientras que la estirpe silvestre mantiene los genes *thn* reprimidos las 15 primeras horas de crecimiento, la estirpe MPO643 presenta una clara desrepresión de los genes *thn* desde el inicio del crecimiento. A partir de las 15 horas, los genes *thn* comienzan a inducirse en la estirpe T1002, alcanzando incluso niveles superiores de inducción que en la estirpe MPO643 al final de la curva. Por tanto, puede confirmarse que SuhB está implicado en la regulación de la represión catabólica de los genes *thn* en TFA.

#### 5.3.4. Complementación funcional de la estirpe MPO643

Para confirmar que el fenotipo observado en el mutante MPO643 ( $\Delta\text{suhB}::\text{kan}$ ;  $\text{thnC}::\text{lacZ}$ ) se debe exclusivamente a la ausencia de SuhB, la estirpe se electroporó con el plásmido pMPO1152 el cual contiene una copia de *suhB* bajo su propio promotor o con el vector vacío pLAFR3. Se analizó la inducción de la fusión *thnC::lacZ* en las estirpes MPO643/pLAFR3 y MPO643/pMPO1152 en células crecidas en medio mínimo con  $\beta$ -HB 40 mM y tetralina procedentes de cultivos en condiciones de no inducción (medio mínimo con  $\beta$ -HB 40 mM). Los resultados de la complementación funcional se representan en la Figura 54.



**Figura 54. Complementación funcional del mutante MPO643 con el vector pMPO1152.** Se muestra la inducción de la fusión *thnC::lacZ* en las estirpes T1002 (TFA; *thnC::lacZ*), MPO643 ( $\Delta\text{suhB}::\text{kan}$ ; *thnC::lacZ*), MPO643/pLAFR3 (vector vacío) y MPO643/pMPO1152 (*suhB*) en medio mínimo con  $\beta$ HB 40 mM y tetralina (condiciones de represión catabólica de los genes *thn*).

Se observa que, tanto en la estirpe MPO643 como en la estirpe MPO643/pLAFR3, los genes *thn* se expresan desde el inicio del crecimiento, mientras que en TFA silvestre (T1002)

## Resultados

y en el mutante complementado (MPO643/pMPO1152), los genes *thn* se mantienen reprimidos durante al menos 10 horas. Por tanto, se demuestra que la desrepresión de los genes *thn* observada en el mutante MPO643 se debe exclusivamente a la ausencia de SuhB, puesto que la estirpe MPO643 complementada con el plásmido pMPO1152 recupera el fenotipo de la estirpe silvestre T1002.

### 5.3.5. Comparación del fenotipo de represión catabólica de las estirpes MPO643 y MPO209-1002

Los mutantes de TFA capaces de expresar los genes *thn* en condiciones de represión catabólica son el mutante de inserción en *phaC* (ver introducción) y el mutante de sustitución  $\Delta$ *suhB::kan* descrito en esta Tesis. Para comparar el fenotipo de ambos mutantes, se realizaron curvas de crecimiento y se midió la actividad  $\beta$ -galactosidasa de la fusión *thnC::lacZ* en las estirpes T1002, MPO643 y MPO209-1002 crecidas en medio mínimo con tetralina y  $\beta$ -HB 40 mM (Figura 55). Los preinóculos con los que se iniciaron las curvas provenían de cultivos en fase exponencial crecidos en medio mínimo con  $\beta$ -HB 40 mM.

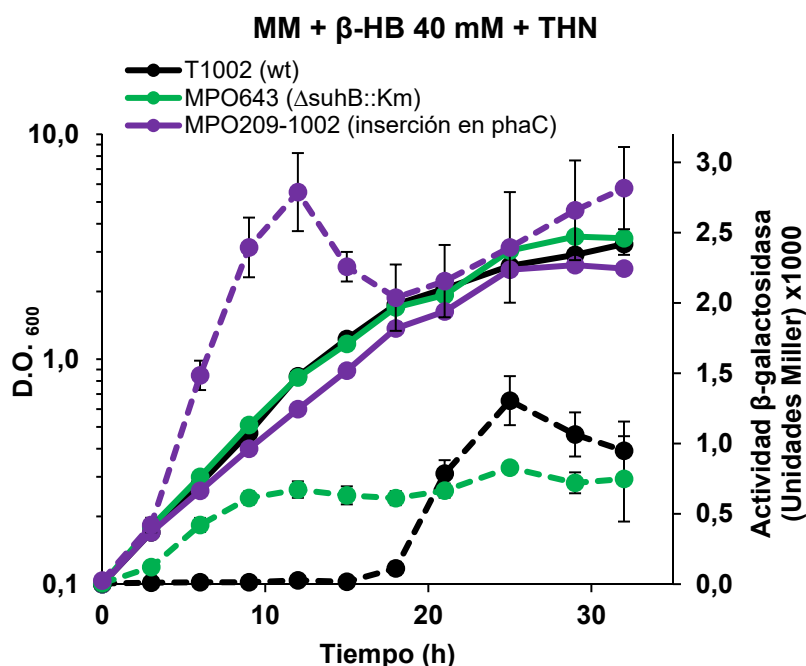


Figura 55. Curva de crecimiento y expresión de la fusión traduccional *thnC::lacZ* en MM con tetralina y  $\beta$ -HB 40 mM en las estirpes T1002 (TFA; *thnC::lacZ*), MPO643 ( $\Delta$ *suhB::kan*; *thnC::lacZ*) y MPO209-1002 (mutante de inserción en *phaC*; *thnC::lacZ*). El crecimiento se representa en línea continua y la actividad  $\beta$ -galactosidasa en línea discontinua.

En la Figura 55 se observa que la fusión *thnC::lacZ* alcanza, desde el inicio del crecimiento, mayores niveles de actividad en la estirpe MPO209-1002 que en la estirpe

MPO643. Además, a partir de las 20 horas de crecimiento, la actividad en la estirpe T1002 es superior a la medida en MPO643 y en la estirpe MPO209-1002 es en torno al doble a la observada en T1002.

### 5.3.6. Caracterización de la expresión de *suhB* en el mutante MPO209 (inserción en *phaC*).

Para comprobar si los dos elementos implicados en la represión catabólica de los genes *thn* estaban relacionados, se decidió estudiar la expresión de *suhB* en el mutante MPO209 (inserción en *phaC*). Para ello, el vector pMPO1151, que contiene la fusión transcripcional *suhB::lacZ*, se integró en el cromosoma de la estirpe MPO209, con lo que se generó la estirpe MPO644 (mutante de inserción en *phaC*, *PsuhB::lacZ*). La expresión del promotor de *suhB* se cuantificó midiendo la actividad  $\beta$ -galactosidasa en células de TFA en fase exponencial de crecimiento en aerobiosis, en un medio mínimo con  $\beta$ -HB 40 mM, con tetralina o con ambas fuentes de carbono. Los cultivos se inocularon a partir de células adaptadas al crecimiento en tetralina, condición de no inducción de *suhB*, tal y como se describe en la Figura 17 del apartado 8.1.1 de materiales y métodos.

Se evaluó el crecimiento de ambas estirpes (Figura 56) y la actividad  $\beta$ -galactosidasa (Figura 57) en las condiciones anteriores.

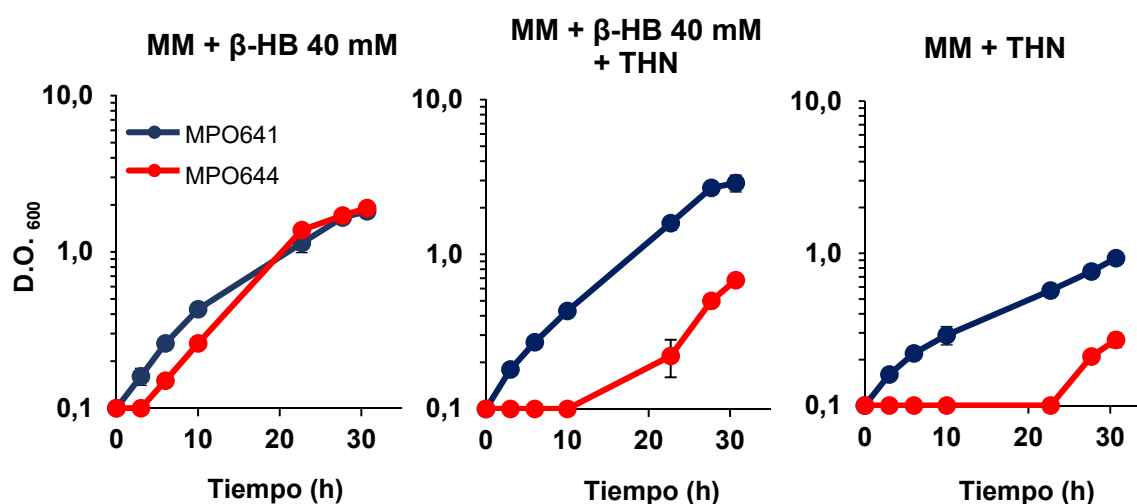


Figura 56. Curvas de crecimiento de las estirpes MPO641 (*PsuhB::lacZ* TFA silvestre) y MPO644 (mutante de inserción en *phaC*; *PsuhB::lacZ*) en MM con  $\beta$ -HB 40 mM, THN o ambas fuentes de carbono.

Se aprecia que en los tres medios estudiados, el crecimiento de la estirpe MPO644 presenta una fase de latencia, lo cual no ocurre en la estirpe MPO641. Además, esta fase es

## Resultados

mayor cuando la tetralina es la única fuente de carbono (Figura 57), mientras que se reduce en medio mínimo con  $\beta$ -HB y tetralina y es mínima cuando el  $\beta$ -HB es la única fuente de carbono.

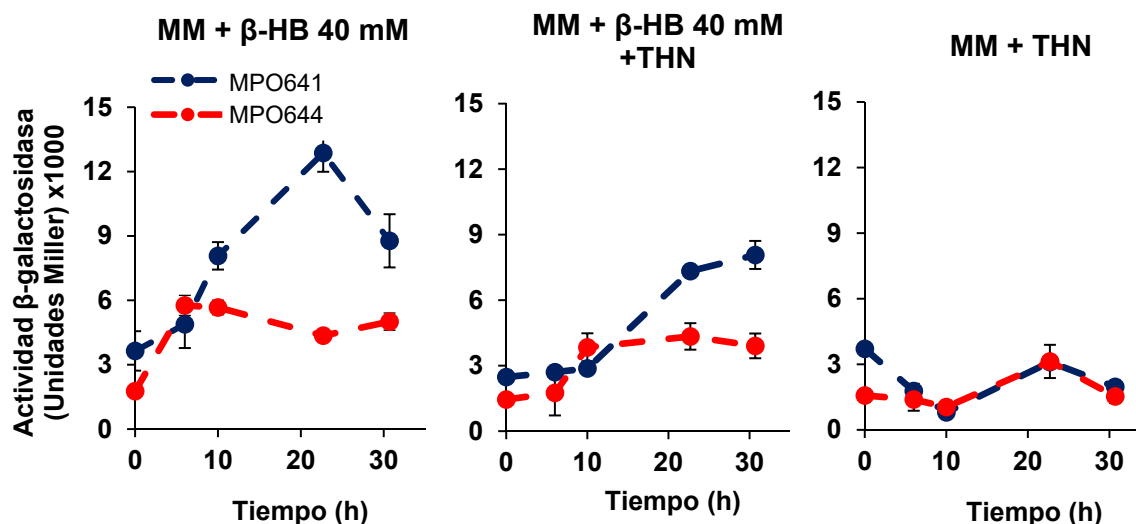


Figura 57. Expresión de la fusión transcripcional *P<sub>suhB</sub>::lacZ* de las estirpes MPO641 (TFA; *P<sub>suhB</sub>::lacZ*) y MPO644 (mutante de inserción en *phaC*; *P<sub>suhB</sub>::lacZ*) en MM con  $\beta$ -HB 40 mM, THN o ambas fuentes de carbono.

Se observa que la expresión de *suhB* en medio mínimo con  $\beta$ -HB 40 mM, con o sin tetralina, es mayor en la estirpe silvestre que en el mutante MPO209, mientras que en tetralina como única fuente de carbono no hay inducción de la expresión de *suhB*. También se observa que, en ambas estirpes, la expresión de *SuhB* es mayor cuando la única fuente de carbono es  $\beta$ -HB 40 mM. Estos resultados corroboran que la presencia de tetralina ejerce un efecto negativo en la expresión de *suhB* y que en un mutante en *phaC*, con una deficiencia en la producción de PHB, la expresión de *suhB* es menor que en TFA silvestre. Por tanto, la expresión de *suhB* de forma directa o indirecta, está influenciada por la acumulación de gránulos de PHB en las células de TFA.

### 5.3.7. Búsqueda de posibles dianas del pequeño ARN *SuhB*

Ante la evidente implicación de *SuhB* en la regulación del fenómeno de represión catabólica de los genes *thn* en TFA, se decidió investigar sus posibles dianas para establecer el mecanismo molecular de acción. Uno de los mecanismos habituales por los cuales los pequeños ARNs que actúan en *trans* ejercen su acción reguladora consiste en su unión al extremo 5' del ARNm diana bloqueando, en la mayoría de los casos, el sitio de unión al ribosoma (Waters y Storz, 2009) impidiendo así la traducción del mismo.



La predicción de posibles dianas de SuhB se realizó *in silico* mediante el programa IntaRNA (Busch y col., 2008). Ante la posibilidad de que la interacción ocurriese en el extremo 5' de los ARNm, la búsqueda se realizó en la región comprendida entre las 150 pb aguas arriba y 150 pb aguas abajo del inicio de la traducción de los genes codificantes para proteínas de TFA (ver materiales y métodos).

En la Tabla 22 se muestran las dianas de SuhB predichas por IntaRNA, ordenadas de menor a mayor energía de interacción.

**Tabla 22. Posibles dianas de SuhB predichas por IntaRNA.** Sólo se muestran las dianas cuya energía de interacción con SuhB es menor de -15 kcal/mol.

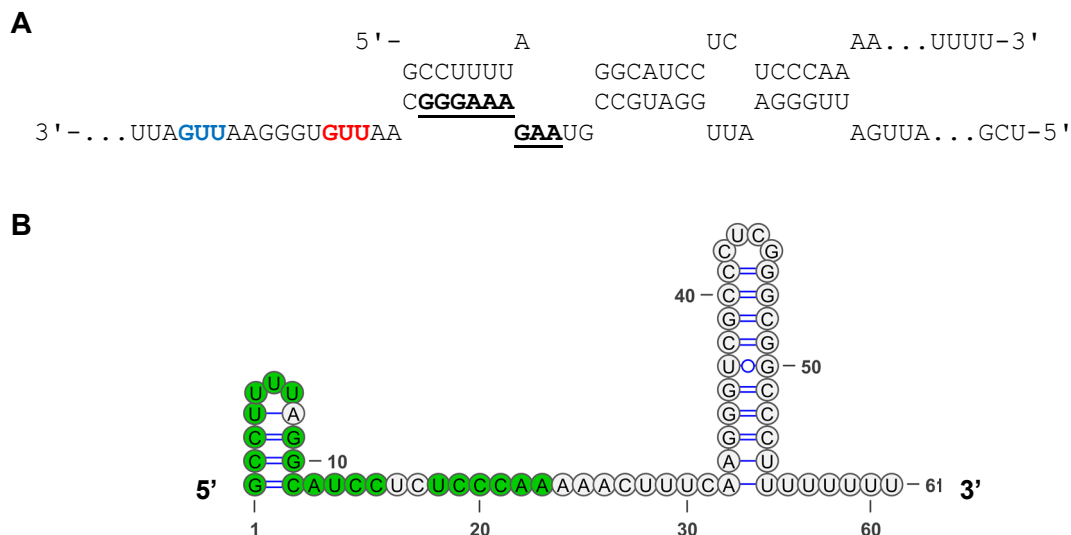
Posible diana	Energía de interacción (kcal/mol)	Descripción
SGRAN_0863	-20,16	Proteína portal de fago
SGRAN_2804 ( <i>thnR</i> )	-17,34	Activador transcripcional tipo LysR
SGRAN_1169	-16,45	Proteína conservada
SGRAN_3690	-16,22	Proteína no caracterizada
SGRAN_1703	-16,20	Proteína no caracterizada
SGRAN_0604	-15,84	Proteína de membrana
SGRAN_2011	-15,75	Proteína facilitadora de transporte
SGRAN_0531	-15,56	Proteína no caracterizada
SGRAN_2891	-15,29	Posible integrasa/recombinasa de fago
SGRAN_1051	-15,07	Receptor dependiente de TonB

Entre las posibles dianas se encuentra el gen *thnR* (regulador transcripcional de los genes *thn*), con una energía de interacción de -17,34 kcal/mol, la segunda más favorable de todas las dianas detectadas.

Junto con la información de la Tabla 22, el resultado de la herramienta IntaRNA también muestra las bases que hibridan entre el pequeño ARN y la diana predicha. En el caso de *thnR*, la interacción se predice en el extremo 5' del transcrito, bloqueando probablemente la secuencia Shine-Dalgarno (SD) de unión al ribosoma (Figura 58A).

Sólo la región 5' de SuhB se une, de manera imperfecta, al ARN mensajero de *thnR* (Figura 58A y 58B). Este tipo de unión con la diana es característica de los pequeños ARNs reguladores que actúan en *trans*, como ya se ha mencionado anteriormente.

## Resultados



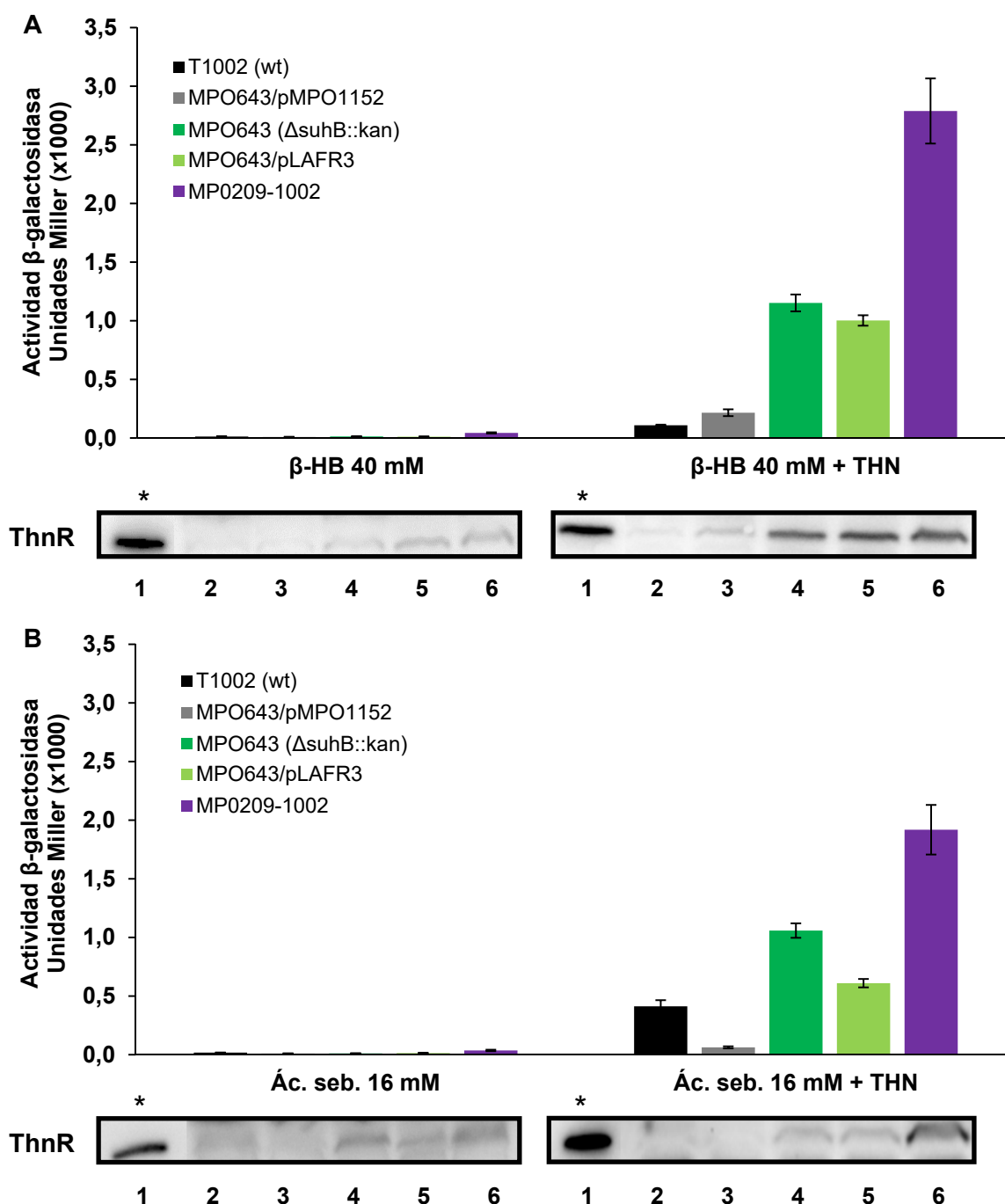
**Figura 58. Representación de la interacción entre SuhB y la región 5'UTR del ARNm de *thnR* predicha por IntaRNA.** (A) Se muestra la unión entre la secuencia de SuhB (5' → 3') predicha por Infernal 1.1 y el extremo 5' del ARNm de *thnR* (3' → 5'). En rojo se indica el codón de inicio de *thnR* detectado por Prodigal en el proceso de anotación, en azul el validado experimentalmente en la Tesis de Martínez-Pérez (2008) y en negrita y subrayada la posible región Shine-Dalgarno de *thnR*. (B) Representación de la estructura secundaria de SuhB, se marcan en verde las bases con las cuales interacciona con el extremo 5' del ARNm de *thnR*.

### 5.3.8. Estudio de la regulación de la expresión de *thnR* mediada por SuhB

Con el fin de analizar la cantidad de ThnR en las estirpes T1002 (*thnC::lacZ*) y MPO643 ( $\Delta$ *suhB::kan*; *thnC::lacZ*), se realizaron experimentos de *Western Blot* con un anticuerpo policlonal anti-ThnR utilizando extractos de células procedentes de cultivos en condiciones de no inducción de los genes *thn* (medio mínimo con  $\beta$ -HB 40 mM o con ácido sebácico 16 mM), crecidas durante 10 horas en condiciones de represión catabólica (medio mínimo con  $\beta$ -HB 40 mM o ácido sebácico 16 mM, y tetralina) o crecidas en condiciones de no inducción (medio mínimo con  $\beta$ -HB 40 mM o ácido sebácico 16 mM). Además, se incorporaron al análisis las estirpes MPO643/pMPO1152 ( $\Delta$ *suhB* complementado) y MPO643/pLAFR3 ( $\Delta$ *suhB*, vector vacío) y MPO209-1002 (mutante de inserción en *phaC*; *thnC::lacZ*)

Como se observa en la Figura 59, en las células crecidas en condiciones de no inducción (crecimiento en medio mínimo con  $\beta$ -HB, A, o ácido sebácico, B), la cantidad de ThnR es mayor en las estirpes MPO643 (carril 4), MPO643/pLAFR3 (carril 5) y MPO209-1002 (carril 6) que en las estirpes T1002 (carril 2) y MPO643/pMPO1152 (carril 3). Esta diferencia aumenta en las células crecidas en condiciones de represión catabólica (mismos carriles que antes pero en medio mínimo con  $\beta$ -HB y tetralina, A, o ácido sebácico y tetralina, B). La cantidad de

ThnR producida correlaciona positivamente con la expresión de la fusión *thnC::lacZ* en las condiciones de represión catabólica (gráficas de la Figura 59).



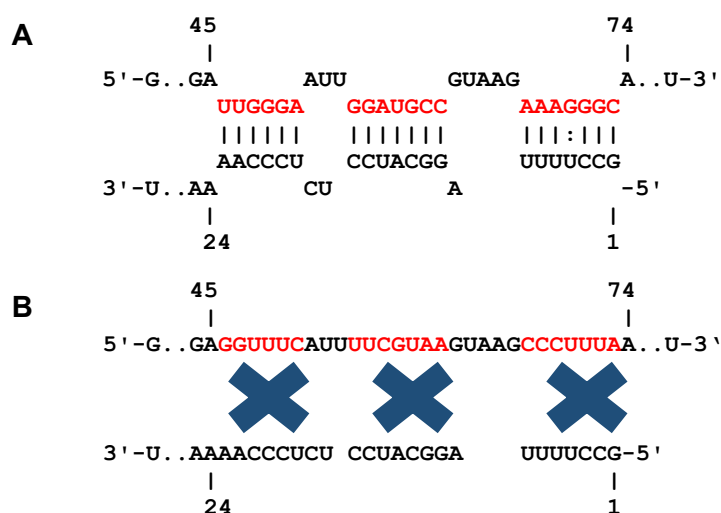
**Figura 59. Detección de ThnR por Western Blot y niveles de expresión de la fusión *thnC::lacZ* a las 10 horas de inducción.** (A) Medidas en medio mínimo con  $\beta$ -HB 40 mM y medio mínimo con  $\beta$ -HB 40 mM y tetralina. (B) Medidas en medio mínimo con ácido sebáico 16 mM y medio mínimo con ácido sebáico 16 mM y tetralina. El asterisco indica la señal de ThnR purificada (1). El resto de bandas corresponden a la detección de ThnR en las mismas estirpes en las que se midió actividad  $\beta$ -galactosidasa, siendo (2) T1002 (silvestre), (3) MPO643/pMPO1152 ( $\Delta$ suhB::kan complementado), (4) MPO643 ( $\Delta$ suhB::kan), (5) MPO643/pLAFR3 ( $\Delta$ suhB::kan) y vector pLAFR3 vacío y (6) MPO209-1002 (mutante de inserción en *phaC*; *thnC::lacZ*).

## Resultados

En el Anexo VI se muestra la cuantificación normalizada de la señal de ThnR en los distintos *Western Blot*, la cual se realizó como se describe en Gürtler y col. (2013) utilizando el programa ImageLab (BioRad)

### 5.3.9. Validación *in vitro* de la interacción entre SuhB y la región 5'UTR de *thnR*

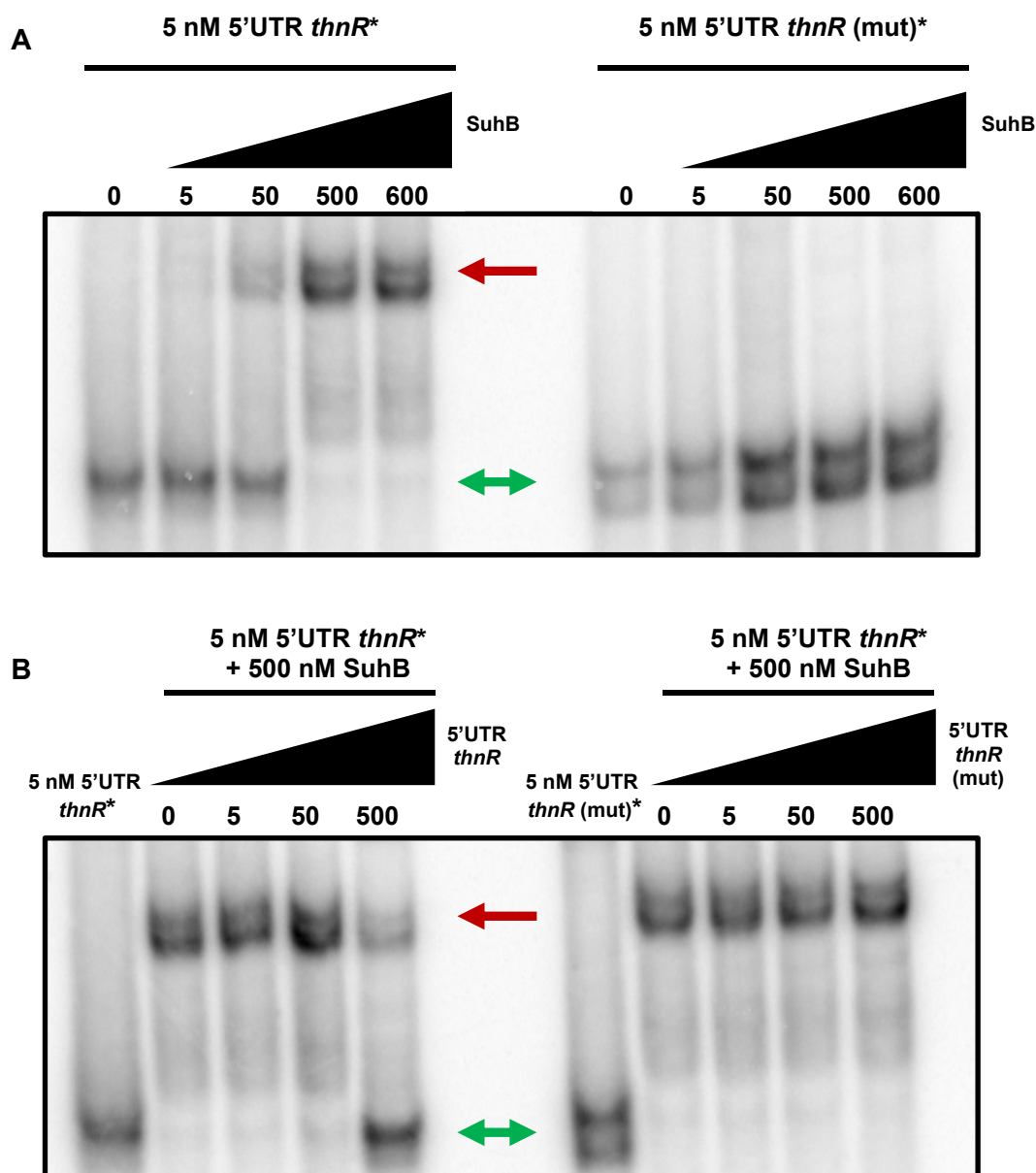
La predicción de la interacción entre SuhB y la región 5'UTR de *thnR* obtenida mediante la herramienta IntaRNA junto con la mayor cantidad de ThnR detectada en MPO643 ( $\Delta$ *suhB::kan*) con respecto a T1002 (silvestre) por *Western Blot*, sugieren que la interacción entre ambas moléculas de ARN bloquea la traducción del regulador. Para comprobar que dicha interacción ocurre *in vitro*, se realizaron ensayos de retardo de movilidad en gel (EMSA). Los transcritos se obtuvieron mediante transcripción *in vitro* con la ARN polimerasa de T7, usando como molde fragmentos lineales de los plásmidos pMPO1155, pMPO1156 y pMPO1157 que contenían las secuencias de *suhB*, la región 5'UTR de *thnR* y una versión de dicha región con los sitios de interacción con SuhB mutados (Figura 60).



**Figura 60. Interacción predicha por IntaRNA entre SuhB y las versiones silvestre y mutante de la región 5'UTR de *thnR*.** (A) 5'UTR *thnR* silvestre (5' → 3') y SuhB (3' → 5'). (B) 5' UTR *thnR* (5' → 3') con los sitios de interacción sustituidos, siendo los cambios U ↔ G y C ↔ A y SuhB (3' → 5'). Las equis azules denotan que con la versión mutada de *thnR* no se detecta interacción entre ambas moléculas a través de la herramienta IntaRNA.

El gen *suhB* se transcribió desde el inicio de la transcripción detectado en los experimentos de dRNA-seq hasta la última timina de la sucesión de timinas que se encuentra tras el terminador independiente de Rho. La región 5'UTR de *thnR*, silvestre y mutada, se transcribió desde el +1 descrito en López y col. (2009) hasta la base +112. Para los ensayos de retardo, se marcó radiativamente el extremo 5' de los transcritos silvestre y mutante de la región 5'UTR de *thnR* y se incubaron con concentraciones crecientes de SuhB. En la Figura 61 se muestran los distintos ensayos de retardo de movilidad en gel realizados para confirmar

la interacción entre el extremo 5'UTR de *thnR* y SuhB. En todos ellos, la banda que se observa en la parte inferior corresponde con el transcrito de la región 5'UTR de *thnR* (silvestre o mutante) marcado radiactivamente. La banda que aparece en la parte superior de los geles corresponde al complejo formado por la unión de los transcritos de *thnR* y SuhB (Figura 61).



**Figura 61. Ensayos de retardo de movilidad en gel del transcrito de la región 5'UTR de *thnR* con SuhB.** Los asteriscos indican que la molécula está marcada radiactivamente. (A) Ensayo de retardo de 5 nM de la región 5'UTR de *thnR* marcado (izquierda) o la versión mutada (derecha) con concentraciones crecientes de SuhB (nM indicados en la Figura). (B) Ensayo de competencia entre 5 nM 5'UTR *thnR* marcado y sin marcar (concentraciones en nM indicadas en la Figura) por la unión con SuhB (izquierda) o el mismo ensayo pero aumentando la concentración de la versión mutada de 5'UTR *thnR* sin marcar (derecha). Las flechas rojas indican el complejo formado y la verde el transcrito de la región 5'UTR de *thnR* libre (silvestre o mutante).

## Resultados

En la Figura 61A se aprecia cómo la formación de dicho complejo ocurre cuando el transcrito de *thnR* marcado se incuba con el transcrito de SuhB, siendo más intensa la señal radiactiva al aumentar la concentración del pequeño ARN. Por el contrario, el complejo no aparece cuando el transcrito marcado es la versión mutada de *thnR*, a pesar de incubarlo con concentraciones crecientes de SuhB.

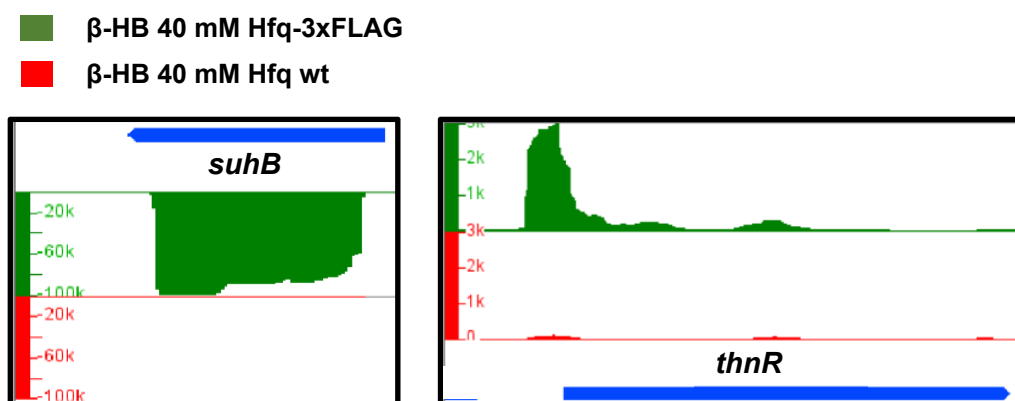
En los ensayos de retardo de la Figura 61B, el complejo formado por el transcrito marcado de *thnR* silvestre y SuhB se pierde cuando se añaden 500 nM del transcrito de *thnR* silvestre sin marcar. Sin embargo, ese mismo complejo no desaparece cuando se incuba con concentraciones crecientes del transcrito de la versión mutada de *thnR*.

Por tanto, los resultados de los ensayos de retardo en gel confirman que la interacción entre la región 5'UTR de *thnR* y SuhB es específica, puesto que SuhB no es capaz de unirse a la versión mutada. Además, se demuestra que la modificación de los sitios de interacción identificados no permite la unión entre ambas moléculas de ARN, por lo que constituye la región de reconocimiento de la diana.

### 5.3.10. Análisis de la implicación de Hfq en la interacción entre *thnR* y SuhB

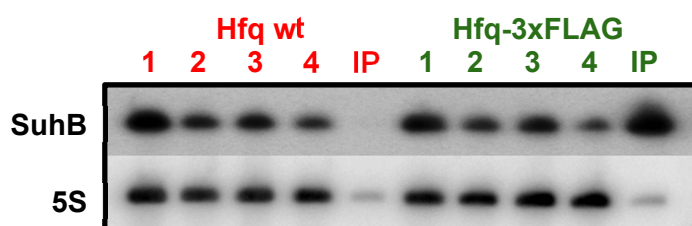
Como se ha mencionado en la introducción, los sRNAs reguladores codificados en *trans*, al contrario de lo que ocurre con los antisentido codificados en *cis*, suelen presentar una limitada complementariedad de bases con sus dianas. En muchos casos, para facilitar la interacción RNAs-ARNm en la regulación *trans*, se requiere de la intervención de la chaperona de ARN conocida como Hfq (Waters y Storz, 2009).

En una primera aproximación para elucidar si los ARNs correspondientes a SuhB y la región 5'UTR de *thnR* podrían unirse a la chaperona Hfq, se realizó un ensayo de co-inmunoprecipitación utilizando la estirpe MPO501 (Hfq-3xFLAG) y TFA silvestre como control, tal y como se describe en Sittka y col., (2008). Las proteínas presentes en extractos de ambas estirpes crecidas hasta fase exponencial en medio mínimo con  $\beta$ -HB 40 mM y los ARNs unidos a ellas se precipitaron con un anticuerpo anti-FLAG. El ARN co-inmunoprecipitado en ambos extractos se retrotranscribió a ADN de doble cadena y se secuenció. En la Figura 62 se muestra el resultado de la co-inmunoprecipitación de SuhB y de la región 5'UTR de *thnR* con Hfq-3xFLAG.



**Figura 62. Co-inmunoprecipitación (CoIP) de SuhB y *thnR* con Hfq-3xFLAG.** Visualización de la señal del RNA-seq de la CoIP en IGB, tanto para *suhB* como para la región 5'UTR de *thnR*, en la muestra con Hfq silvestre (wt) y la muestra Hfq-3xFLAG.

En cada etapa de la co-inmunoprecipitación se tomaron muestras que se sometieron a análisis por *Northern Blot* para la detección de SuhB. Como se observa en la Figura 63, SuhB está presente en la estirpe silvestre y en la estirpe con Hfq-3xFLAG en todos los pasos del proceso salvo en la co-inmunoprecipitación, donde no se detecta en la estirpe silvestre y sí en la estirpe con Hfq-3xFLAG.



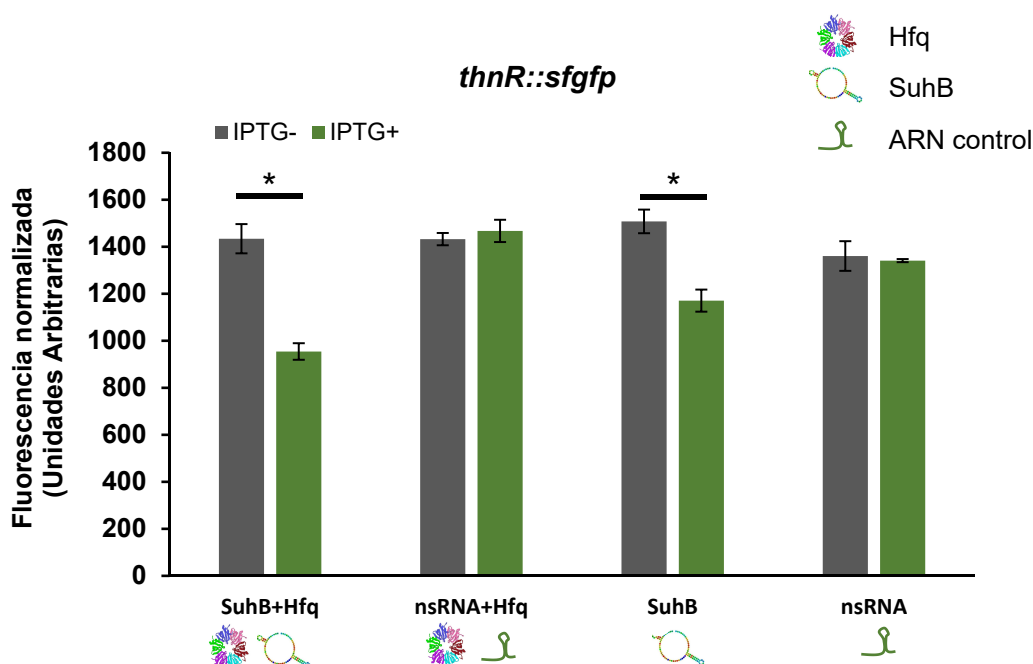
**Figura 63. *Northern Blot* con la sonda complementaria a SuhB en las distintas etapas de la co-inmunoprecipitación.** Las muestras de ARN se extrajeron a lo largo del proceso de la co-inmunoprecipitación: (1) Células, (2) lisado, (3) después del anticuerpo anti-FLAG, (4) sobrenadante tras la adición de proteína A sefarosa, (IP) muestra de CoIP.

Para confirmar el resultado de la co-inmunoprecipitación, se estudió la interacción *in vivo* entre la región 5'UTR de *thnR*, SuhB y la proteína Hfq en un organismo heterólogo (*E. coli* TOP10F') a través de un sistema de tres plásmidos basado en el gen reportero *sfGFP* (ver apartado 9.1.1 de materiales y métodos). La región 5'UTR completa de *thnR*, desde su inicio de transcripción, junto con 81 nucleótidos de la región codificante (27 aa, atendiendo a la anotación realizada en esta Tesis) se fusionaron traduccionalmente a la proteína sfGFP. El plásmido pXG10-SF que contiene la fusión *thnR::sfgfp* (pMPO1154) se introdujo en células de *E. coli* TOP10F' que posteriormente se transformaron con las siguientes combinaciones de plásmidos; pMPO1153 ( $P_L$ -*suhB*), o pJV300 (control) y con pMPO1819 ( $P_{tac}$ -*hfq*) o pIZ1016

## Resultados

(control). Como control interno se construyeron las mismas combinaciones con el plásmido pXG0, el cual no contiene sfGFP.

La estirpe de *E. coli* TOP10F', con las distintas combinaciones de plásmidos descritas en el párrafo anterior, se cultivó en medio LB hasta la saturación para diluirla después 100 veces en LB y LB más 1 mM de IPTG para la inducción de los pequeños ARNs y de *hfq*. Tras 6 horas de crecimiento, los cultivos se diluyeron 2 veces ( $F_d=2$ ), se lavaron con PBS y se les midió la D.O.<sub>600</sub> y la fluorescencia a través de un fluorímetro. Una vez obtenidos dichos valores, se normalizó la fluorescencia emitida por cada una de las combinaciones de plásmidos, con y sin IPTG, tal y como se detalla en el apartado 9.1.1.2 de materiales y métodos (Figura 64).



**Figura 64. Fluorescencia de la fusión traduccional *thnR::sfGFP* normalizada con respecto al control pXG0.**

Los símbolos indican la presencia de los distintos elementos del sistema: la proteína Hfq, el pequeño ARN SuhB y el ARN control. El asterisco denota un  $p$ -value  $< 0,005$  en el test T-student realizado entre las muestras.

Como se observa en la Figura 64, la fluorescencia procedente de la fusión *thnR::sfGFP* disminuye cuando se expresa *suhB* pero no lo hace cuando se expresa el ARN control. Además, esta disminución en la fluorescencia es ligeramente mayor si, además de SuhB, se induce la expresión de *hfq*, mientras que con el ARN control, la fluorescencia no varía aunque se exprese *hfq*. Por tanto, puede afirmarse que la expresión de *suhB* disminuye la traducción de *thnR* y dicha disminución se ve favorecida por Hfq, al interaccionar los tres elementos entre sí, el pequeño ARN SuhB, la región 5'UTR de *thnR* y la chaperona Hfq, para evitar la traducción de ThnR.



## **DISCUSIÓN**

---



## 1. Organización y características del genoma de *S. granuli* TFA

El tamaño definido para el genoma de TFA tras el procedimiento de secuenciación y ensamblado (4,7 Mb) se encuentra en la parte alta del intervalo definido para el género *Sphingopyxis* que, según los genomas secuenciados hasta el momento, varía entre 3 Mb (*S. baekryungensis* DSM 16222) y 4,99 Mb (*S. fribergensis* Kp5.2). Del mismo modo, el contenido en G+C del genoma de TFA (66,2 %) tiene un valor intermedio dentro de los valores definidos para el género *Sphingopyxis*, que oscilan entre el 62,3 % y el 69,2 % (Glaeser y Kämpfer, 2014). Considerando el contenido génico, la proporción de región codificante en el genoma de TFA supone el 89,47 % del total de la secuencia genómica. Este valor es muy cercano al 88 % de media establecido en un estudio de 2679 genomas bacterianos completos extraídos de GenBank (sin considerar ARNs funcionales) (Land y col., 2014; Land y col., 2015).

Si bien los experimentos para la detección de plásmidos y mega-plásmidos en TFA arrojaron un resultado negativo, (Hernández Silva, 2000), tal y como se planteó el cierre del genoma de TFA (Figura 20), el fragmento ctg220-217 de 1,8 Mb podría constituir un replicón independiente, con un tamaño similar al mega-plásmido pSA1 de *Novosphingobium resinovorum* estirpe SA1 (Número de acceso de GenBank CP017076) (Hegedüs y col., 2017). Sin embargo, esta hipótesis ha de descartarse debido a que dicho fragmento contiene el único operón codificante para los ARNr que se ha anotado en el genoma de TFA, por lo que el fragmento tiene que formar parte necesariamente del cromosoma bacteriano, que requiere, por definición, la presencia de al menos un operón ribosómico.

El número de genes codificantes de ARNt en el genoma de TFA asciende a 46. Este número es el mismo que para *S. fribergensis* Kp5.2, similar a los 45 detectados en *S. alaskensis* RB2256 y *S. terrae* NBRC 15098 203-1 y a los 47 de *Sphingopyxis* sp. 113P3 y ligeramente inferior a los 50 y 58 encontrados en los genomas de *S. macrogoltabida* EY-1 y *S. macrogoltabida* 203 respectivamente. A pesar de que es posible que existan ARN transferentes con hasta 61 anticodones diferentes en una bacteria, el estudio llevado a cabo por Land y col. (2014) mostró que el número máximo de ARNt encontrados en más de 20000 genomas bacterianos, de 32000 analizados, fue de solo 47. Por otra parte, la cantidad de genes codificantes de ARNt se ha correlacionado positivamente con la velocidad de crecimiento. En este sentido, un estudio realizado con 102 genomas extraídos de GenBank determinó que las bacterias con un crecimiento rápido tenían una media de 61 ARNt mientras que las de crecimiento lento de 44 ARNt (Rocha, 2004). A su vez, una velocidad de crecimiento reducida se ha atribuido a un estilo de vida oligotrofo, como se ha descrito para *S. alaskensis* RB2256, la cual alcanza una velocidad máxima de crecimiento que no supera las 0,2 h<sup>-1</sup> (Lauro y col., 2009).

## Discusión

Los 46 ARNt encontrados en el genoma de TFA son suficientes para explicar la síntesis de todas las proteínas codificadas en el genoma, bien de forma directa, considerando la regla de tambaleo o con la intervención de nucleósidos modificados en la posición de tambaleo. Tanto el gen *tilS*, responsable de la síntesis de lisidina, como el gen *tadA*, de la síntesis de inosina, se encuentran conservados en las bacterias de la familia *Sphingomonadaceae*. Además, ambos genes se encuentran dentro del genoma núcleo del género *Sphingopyxis* obtenido en esta tesis doctoral. En la mayoría de bacterias y arqueas, un anticodón GAU (ARNt<sup>Ile</sup><sub>GAU</sub>) lee dos de los codones de isoleucina (AUU y AUC), atendiendo a la regla clásica de tambaleo, mientras que el tercer codón, AUA, se lee con la intervención del nucleósido lisidina (ARNt<sup>Ile</sup><sub>LAU</sub>), el cual deriva de la C del anticodón CAU que codifica para metionina (Grosjean y Björk, 2004; Köhrer y col., 2013). Del mismo modo, para la decodificación de los 4 codones de arginina CGN, la mayoría de bacterias presentan solo dos ARNt distintos, uno de ellos porta el anticodón CCG (ARNt<sup>Arg</sup><sub>CCG</sub>) y lee el codón CGG y el otro el anticodón ACG (ARNt<sup>Arg</sup><sub>ACG</sub>) el cual lee el codón CGU y, gracias a la deaminasa TadA que convierte la adenosina en inosina, también lee los codones CGC y CGA (ARNt<sup>Arg</sup><sub>ICG</sub>) (Wolf y col., 2002; Yokobori y col., 2013). Por tanto, aunque la presencia de los genes *tilS* y *tadA* permite reducir el número mínimo de ARNt para leer todos los codones, su conservación entre las *Sphingomonadaceae* no supone ningún rasgo exclusivo o distintivo de oligotrofos, ya que estos genes se encuentran en multitud de bacterias distintas.

Atendiendo a la función de las proteínas, un total de 688 no tienen función asignada incluso tras la revisión manual de la anotación, lo cual se corresponde con un 16,4 % del total. Incluso en el genoma núcleo del género *Sphingopyxis*, las categorías R (función general) y S (función desconocida) son de las más representadas. Esta misma predominancia de proteínas no caracterizadas en el genoma núcleo de las *Sphingopyxis* se había observado previamente en los géneros *Sphingomonas*, *Novosphingobium* y *Sphingobium* al definir grupos de genes que estuviesen sobre-representados en uno de esos géneros frente a los otros dos (Aylward y col., 2013), lo cual sugiere que estas bacterias albergan muchas funciones desconocidas cuyo estudio podría resultar de gran interés para entender mejor su fisiología y capacidades metabólicas. En *Escherichia coli*, la bacteria más estudiada hasta el momento, la proporción de proteínas con función desconocida era de aproximadamente el 10 % en 2007 (Matte y col., 2009).

El origen de replicación del genoma de TFA (OriC) se obtuvo con la herramienta Ori-Finder, la cual tiene en cuenta la presencia de genes que comúnmente se encuentran cerca de los orígenes de replicación, la acumulación de cajas DnaA y la asimetría en la composición de bases del genoma. Según la información de la base de datos DoriC (Gao y col., 2013), el

origen de replicación de la mayoría de las bacterias de la familia *Sphingomonadaceae* se encuentra próximo al gen *hemE*, tal y como ocurre en TFA. Además, en esta tesis se ha demostrado que tanto el contexto génico del OriC como las posibles cajas DnaA y el elemento DUE están conservados en todos los genomas estudiados del género *Sphingopyxis* (Figura 23). La presencia de los genes conservados podría utilizarse para identificar el origen de replicación en la clase  $\alpha$ -proteobacteria como ya se propuso en Brassinga y col., 2001. De acuerdo a lo anterior, el OriC de *S. alaskensis* RB2256, anotado en la base de datos DoriC en una región distinta, debería moverse a la región cercana al gen *hemE*.

Otro método ampliamente extendido para asignar tanto el origen de replicación como la terminación en un genoma bacteriano es el análisis de la desviación GC, definida como el cálculo de Gs sobre Cs ( $G-C/G+C$ ) a lo largo del genoma (Grigoriev, 1998; Lobry, 1996). En el mecanismo de replicación bidireccional, teniendo en cuenta que los nucleótidos sólo pueden ser añadidos al extremo 3', se diferencian dos cadenas, una de las cuales se sintetiza en la misma dirección que el movimiento de la horquilla de replicación, cadena líder, y la otra en la dirección opuesta, cadena retardada. Además, la cadena retardada se sintetiza como pequeños fragmentos (fragmentos de Okazaki) que deben ser unidos posteriormente (Kornberg y Baker, 1992). Esta replicación bidireccional resulta en una presión mutacional distinta entre la cadena líder y retardada, que se refleja en una composición de bases diferente.

Se ha propuesto que uno de los mecanismos responsables de la asimetría en la composición de bases sea la desaminación de la citosina, la cual ocurre con mayor frecuencia en ADN de cadena sencilla que de cadena doble (Frank y Lobry, 1999). Dicha desaminación transforma la C en U, lo cual puede ser corregido por la enzima uracil-ADN glicosilasa (Coulondre y col., 1978), pero si la citosina está metilada, la desaminación produce T, lo cual no puede corregirse por la enzima anterior. Debido al mecanismo de replicación, la cadena líder está más expuesta al estado de cadena sencilla (para servir como molde para la síntesis de la cadena retardada, que ocurre de forma lenta y discontinua), y las mutaciones de sustitución de C por T son más frecuentes que en la cadena retardada. Como consecuencia, la cadena líder está enriquecida en G frente a C en comparación con la cadena retardada, tal y como se observa en el genoma de TFA (Figura 24, nótese que sólo se representa una de las cadenas del genoma en sentido 5'  $\rightarrow$  3'. Al considerar la horquilla de replicación desde el OriC, la parte de la derecha constituye la cadena líder, molde de la retardada, y la de la izquierda la cadena retardada, molde de la líder).

Sin embargo, el OriC detectado en el genoma de TFA se encuentra desplazado unas 300 kb de la inflexión en la desviación GC. El mismo fenómeno se ha detectado en otras bacterias

## Discusión

como *Helicobacter pylori* 26995 (Mackiewicz y col., 2004), aunque en este caso, el predicho solo estaba desplazado 2,5 kb del real. Cambios locales en la desviación GC podrían deberse a reorganizaciones génicas e incorporaciones recientes de ADN externo (Ioannidis y col., 2007; Eppinger y col., 2004), lo cual dificultaría la detección del OriC por esta metodología. Si esos cambios ocurren alrededor del OriC y no ha pasado suficiente tiempo para que la presión mutacional vuelva a generar una desviación clara entre la cadena líder y retardada, puede obtenerse una predicción errónea. En el genoma de TFA se han detectado numerosas islas genómicas cercanas al origen de replicación (Figura 25), justo en la zona de desfase entre el OriC y la inflexión en la desviación GC, lo que ha podido generar la alteración en la desviación GC observada en dicha zona. Por tanto, en algunos genomas, los cambios de inflexión en la desviación GC pueden conducir a error en la localización del origen de replicación y se requiere tener en cuenta otros parámetros como los que se consideran en la herramienta Ori-Finder.

Mediante la comparación de la secuencia del genoma de TFA con genomas cercanos filogenéticamente y la detección de islas genómicas en su secuencia, se han delimitado distintas regiones que podrían corresponder con zonas que la bacteria ha adquirido a través de transferencia horizontal, tal y como se ha mencionado anteriormente para la secuencia cercana al OriC. Todas estas regiones presentan un contenido en G+C por debajo de la media del genoma (Figura 25) y muchas de ellas no presentan similitud ni siquiera con la estirpe ku-sg de la misma especie que TFA, *Sphingopyxis granuli*. Los genomas de TFA y *S. granuli* ku-sg son casi idénticos salvo en las posibles islas genómicas. Curiosamente, en algunas de estas islas del genoma de TFA se encuentran fragmentos con una identidad próxima al 100 %, con el genoma de otras bacterias. Muchas de estas zonas con alta similitud contienen genes relacionados con la transferencia de plásmidos por conjugación (genes *tra/trb*) lo cual sugiere que podría tratarse de plásmidos que se han integrado y mantenido en el genoma de TFA. Adicionalmente, se han identificado genes *rep* y *par*, implicados en la replicación y partición de plásmidos, respectivamente, tanto agrupados en islas genómicas como dispersos a lo largo del genoma. Todas estas proteínas Rep y Par de TFA parecen estar emparentadas filogenéticamente. En concreto, todas las proteínas RepA anotadas en el genoma de TFA, encargadas de la regulación del número de copias de los plásmidos, se engloban dentro de la super-familia RPA (Pfam 10134), mientras que en los plásmidos del grupo de las esfingomonas se han detectado proteínas Rep pertenecientes a cuatro super-familias distintas, que son RepA\_C (Pfam 04796), Rep\_3 (Pfam 01051), HTH-36 (Pfam 13730) y RPA. Por otra parte, todas las proteínas ParA de TFA, que aseguran la correcta distribución de las copias de los plásmidos a las células hijas, se adscriben a la super-familia CbiA (Pfam 01656).

Se ha correlacionado la presencia de un tipo de proteína RepA con un determinado grupo de incompatibilidad de manera que dos plásmidos con proteínas RepA pertenecientes a la misma super-familia serían del mismo grupo de incompatibilidad y, como consecuencia, no podrían coexistir como replicones independientes en una misma bacteria (Stolz, 2014). Experimentalmente se ha demostrado que la transferencia del plásmido pNL1 de *Sphingomonas aromaticivorans* F199 a *Sphingomonas sp. HH69*, cuyo mega-plásmido de 240 kb pertenece al mismo grupo de compatibilidad que pNL1 (Basta y col., 2005), genera una reorganización génica en condiciones selectivas para ambos plásmidos. Dicha reorganización puede consistir en la generación de un único plásmido o en la integración de parte de los plásmidos en el cromosoma de la bacteria receptora (Basta y col., 2004). Esto demuestra que, bajo presión selectiva, los plásmidos incompatibles pueden reorganizarse de tal forma que los genes responsables de una mejor adaptación al medio se incorporen a otro replicón, como puede ser el cromosoma. En base a estos precedentes y a que todas las proteínas RepA detectadas en TFA pertenecen a la misma superfamilia (RPA), el genoma de esta bacteria podría haber sufrido reorganizaciones del mismo tipo al haber adquirido plásmidos incompatibles cuya presencia hubiera sido seleccionada en condiciones ambientales.

Además de genes característicos de plásmidos, en el genoma de TFA se han anotado numerosos genes relacionados con elementos móviles (46 codificantes para transposasas y 30 para integrasas), así como tres regiones que podrían ser profagos. La presencia de estos elementos resulta muy común en los genomas de las esfingomonas, tal y como se concluyó en un estudio donde se compararon 26 genomas de este grupo de bacterias (Aylward y col., 2013).

La abundante presencia de genes relacionados con elementos móviles y la posible reorganización génica forzada por la coexistencia de plásmidos del mismo grupo de incompatibilidad, han podido constituir las fuerzas prominentes con las cuales se ha obtenido la organización actual del genoma de TFA. De hecho, este tipo de elementos ya se han postulado como claves en la evolución y constitución de los genomas de las esfingomonas, pudiendo ser los responsables del frecuente intercambio de capacidades degradadoras en este grupo de bacterias (Aylward y col., 2013).

Un número reducido de operones ribosómicos (sólo se detecta uno en TFA) se asocia comúnmente con bacterias que habitan en nichos especializados y que no son capaces de responder rápidamente a cambios en las condiciones ambientales, así como a bacterias de crecimiento lento que viven en condiciones de oligotrofía (baja concentración de nutrientes), como es el caso de *Sphingopyxis alaskensis* especie RB2256 (Fegatella y col., 1998; Klappenbach y col., 2000). En especies del grupo de las esfingomonas, se han detectado

entre 1 y 4 copias (Aylward y col., 2013). Además, dentro del genoma núcleo del género *Sphingopyxis*, las categorías COG de movilidad celular (N) y de mecanismos de defensa (V) presentan una baja representación, mientras que la de transporte y metabolismo de lípidos (I) es una de las más representadas. Este resultado también concuerda con las características genómicas descritas para las bacterias con un estilo de vida oligotrofo (Lauro y col., 2009).

### 1.1. Adscripción filogenética de TFA a la especie *S. granuli*, estilo de vida y nicho ecológico

La comparación de la secuencia del genoma de TFA frente a los genomas de otras bacterias de la clase  $\alpha$ -proteobacteria (Figura 25), así como los distintos estudios comparativos y filogenéticos realizados en esta tesis doctoral (Figuras 26, 27, 28 y Tablas 14 y 15), pusieron de manifiesto la cercanía filogenética entre TFA y las estirpes de la especie *Sphingopyxis granuli*. Previamente, TFA se clasificó como *Sphingomonas macrogoltabida* mediante el alineamiento de la secuencia parcial del gen codificante para el ARNr 16S, lo cual demuestra que la adscripción basada en la secuencia parcial de este gen puede dar lugar a identificaciones erróneas si, además, en las bases de datos no hay suficiente representación del género y la especie bacteriana en estudio.

Los genomas de las estirpes *Sphingopyxis granuli* TFA y *Sphingopyxis granuli* ku-sg son muy parecidos a nivel de secuencia de nucleótidos, compartiendo grandes regiones con una similitud cercana al 100 %. La secuencia del genoma de la estirpe *S. granuli* ku-sg se obtuvo, de forma indirecta, en un proyecto de secuenciación del genoma de una planta de arroz (*Oryza sativa* variedad indica cultivar RP Bio-226). En el ensamblaje de las lecturas de la secuenciación de RP Bio-226 frente a 2101 genomas bacterianos de referencia se identificó tanto a *S. granuli* como a *P. aeruginosa* como bacterias endofíticas de la planta (Battu y col., 2017a; Battu y col., 2017b). Como se mencionó en la introducción, las bacterias del grupo esfingomonas (compuesto por los géneros *Sphingomonas*, *Sphingopyxis*, *Sphingobium* y *Novosphingobium*) son ubicuas y algunas se han aislado incluso de animales o plantas. En el caso de las esfingomonas endofíticas, además de *S. granuli* ku-sg se han aislado otras como *Novosphingobium oryzae* estirpe ZYY112 (Zhang y col., 2016), que también es endofítica de arroz, o *Sphingomonas* SaMR12, la cual se aisló de *Sedum alfredii*, una planta que se encuentra en algunas áreas de Asia (Chen y col., 2014).

Se ha demostrado que las bacterias endofíticas promueven el crecimiento de las plantas que colonizan a través de la producción de compuestos que estimulan el crecimiento, como el ácido indolacético (AIA), cuya producción se ha validado en bacterias endofíticas como *Azoarcus* sp. CIB (Fernández y col., 2014). Además, en un estudio realizado con bacterias



del filo Proteobacteria se identificaron genes que podrían tener relación con el estilo de vida endofítico (Ali y col., 2014). Entre ellos se encuentran distintos genes que codifican para proteínas transportadoras, sistemas de secreción tipo VI, bombas de eflujo RND (acrónimo de *Resistance-Nodulation-cell Division*), distintas hidrolasas, proteínas de la super-familia cupina, o genes relacionados con detoxificación, como los de la síntesis y la reducción de glutatión.

Aunque las bombas RND pueden tener diversas funciones (Álvarez-Ortega y col., 2013), su presencia se ha relacionado con la naturaleza endofítica de la bacteria *Enterobacter* sp. 638 (Taghavi y col., 2010). En concreto, uno de estos transportadores RND (AcrAB) es homólogo al utilizado por *Erwinia amylovora* para exportar el metabolito secundario fitoalexina, necesario para poder colonizar la planta (Burse y col., 2004). En el genoma de TFA se han anotado varias agrupaciones de genes que podrían codificar para los componentes de un sistema de bomba RND: una proteína de membrana externa, una proteína de fusión de membrana periplásmica y el transportador RND. Sin embargo, solo uno de los sistemas transportadores es homólogo a AcrAB de *Erwinia amylovora*, cuya proteína de membrana externa no ha sido identificada (Burse y col., 2004). AcrA, la proteína de fusión de membrana periplásmica, presenta una identidad del 44 % con la proteína codificada por el gen SGRAN\_4082 y AcrB, el transportador de tipo RND, del 53 % con SGRAN\_4083. En TFA, el gen SGRAN\_4084 podría codificar para la proteína de membrana externa de la bomba RND. Debido al parecido con la bomba RND de *E. amylovora*, podría plantearse que en TFA tuviese una función parecida.

Por otra parte, se ha propuesto que las proteínas pertenecientes a la super-familia cupina, de las que TFA tiene 2 representantes anotados (SGRAN\_0738 y SGRAN\_3802), participen en la modificación de los carbohidratos de la pared celular vegetal (Dunwell y col., 2004) y se ha demostrado que los sistemas de secreción tipo VI ejercen un papel positivo en las interacciones planta-microbio (Mattinen y col., 2008). No se han analizado todos los componentes de un sistema de secreción tipo VI en TFA; sin embargo, en el genoma se encuentran anotadas 5 proteínas de secreción tipo VI (SGRAN\_0056, SGRAN\_0184, SGRAN\_0480, SGRAN\_1865 y SGRAN\_4231) que podrían formar parte de estos sistemas.

A su vez, la presencia de glutatión S-transferasas (GST), de las que TFA posee 13, y glutatión peroxidasas, de las que TFA posee 2, es consistente con un posible efecto de respuesta de protección de la bacteria frente al estrés oxidativo creado como consecuencia de la infección a la planta (Ali y col., 2014). El número de GST en TFA es mayor que el de la bacteria endofítica *Enterobacter* sp. 638 (3 GST) y muy cercano al de la bacteria endofítica *Klebsiella pneumoniae* 342 (12 GST) (Taghavi y col., 2010; Fouts y col., 2008). La presencia

de GST también se ha asociado a la degradación de compuestos aromáticos como el tolueno, xilenos, fenoles y la atrazina (Santos y col., 2002; Allocati y col., 2008).

Aunque no puede afirmarse de forma contundente por falta de evidencia experimental, los indicios expuestos anteriormente sugieren que TFA podría colonizar células vegetales y tener un estilo de vida endofítico o asociarse a la rizosfera para proporcionar algún tipo de beneficio a la planta. De hecho, las bacterias del grupo esfingomonas cumplen un papel importante en la descontaminación de suelos a través de plantas (fitorremediación) asistida por microorganismos (revisado en Waigi y col., 2017).

Por otra parte, las bacterias del grupo esfingomonas se han caracterizado como estrictamente aerobias (Takeuchi y col., 2001; Balkwill y col., 2006). Sin embargo, en los genomas de *S. granuli* TFA y *S. granuli* ku-sg se han identificado genes de respiración de nitrato. De hecho, se ha demostrado la capacidad de *S. granuli* TFA para crecer anaeróbicamente con nitrato como último aceptor de electrones (García-Romero y col., 2016). Además, en *S. granuli* Kw07, cuyo genoma no ha sido secuenciado, se ha descrito una reducción de nitrato a nitrito, aunque no se ha demostrado que la bacteria pueda crecer respirando nitrato anaeróbicamente (Kim y col., 2005). Todos estos datos indican que existen excepciones a la regla general y que éstas se circunscriben a cepas de la especie *granuli*. Se ha establecido que la reducción de nitrato en *Pseudomonas fluorescens* confiere una ventaja selectiva en la colonización de la rizosfera (Ghiglione y col., 2002). Por tanto, esta particularidad de la especie *S. granuli* podría estar relacionada con su nicho ecológico, en concreto, con el estilo de vida endofítico o asociado a la rizosfera. En el caso de los cultivos de arroz, los cuales se caracterizan por realizarse en suelos encharcados, se diferencian zonas óxicas y anóxicas (Breidenbach y col., 2016) por lo que bacterias con un metabolismo anaerobio facultativo como TFA podrían tener una ventaja adaptativa frente a microorganismos estrictamente aerobios.

## 2. Estudio de las capacidades metabólicas a través del modelo iIG738

El modelo metabólico iIG738, obtenido tras la conversión matemática de la reconstrucción del metabolismo de TFA, ha sido capaz de simular capacidades metabólicas que se han estudiado experimentalmente en TFA, como la respiración anaerobia de nitrato o el consumo de distintos ácidos grasos y aminoácidos. Además, ha permitido conocer nuevas fuentes de carbono que TFA es capaz de utilizar.

Previamente a la generación del modelo metabólico iIG738, las únicas fuentes de carbono y energía que se habían utilizado para el crecimiento de TFA eran  $\beta$ -HB, prolina, glutamina,

acetato y ácidos grasos de cadena media-larga como el ácido sebácico (10 C), el ácido mirístico (14 C) o el ácido tetradecanodioico (14 C). La construcción y validación del modelo metabólico puso de manifiesto la utilización por parte de TFA de nuevos compuestos como únicas fuentes de carbono y energía, tales como otros ácidos grasos, nuevos aminoácidos, incluyendo los aromáticos, y el ácido láctico (ver Tabla 17). Sin embargo, entre estos metabolitos no se encuentran azúcares, a pesar de que el modelo predice crecimiento si se incorporan las reacciones de transporte de estos metabolitos.

En el genoma de TFA no se han identificado transportadores de azúcares pero sí que ha sido posible reconstruir completamente las rutas de Embden-Meyerhof-Parnas (Glucolisis) y de Entner-Doudoroff, por las cuales se metabolizarían estos azúcares hasta llegar al ciclo de los ácidos tricarboxílicos (TCA). Por tanto, atendiendo a la reconstrucción metabólica, es posible plantear que la expresión heteróloga de transportadores de glucosa o fructosa en la estirpe silvestre de TFA pueda permitir el crecimiento en estos monosacáridos como única fuente de carbono y energía, aunque se requiere demostración experimental. De hecho, en *Shewanella oneidensis*, una bacteria marina que no puede utilizar glucosa debido a la falta de transportador y del primer paso de fosforilación, la introducción del gen del transportador de glucosa (*gluP*) y de una hexoquinasa (*glk*) de *Zymomonas mirabilis* permitió el crecimiento de esta bacteria en glucosa (Choi y col., 2014).

La imposibilidad de utilizar glucosa como fuente de carbono y energía no es un rasgo exclusivo de TFA dentro del género *Sphingopyxis* aunque, hasta el momento, sí lo es de la especie *S. granuli*. Se ha demostrado experimentalmente que *S. granuli* Kw07 no es capaz de asimilar la D-glucosa, mientras que el resto de *Sphingopyxis* analizadas sí que mostraron la habilidad de utilizarla (*Sphingopyxis alaskensis* DSM 13593, *Sphingopyxis chilensis* DSM 14889, *Sphingopyxis macroglabida* DSM 8826 y *Sphingopyxis witflariensis* DSM 14551) (Kim y col., 2005). Por otra parte, aunque en un análisis de las fuentes de carbono y nitrógeno utilizadas por *S. alaskensis* RB2256 (cuyo genoma está secuenciado), se detectó que esta bacteria es capaz de utilizar glucosa como única fuente de carbono y energía, en su genoma no se identificaron los genes codificantes del sistema PTS (sistema fosfotransferasa) de transporte y fosforilación de la glucosa (Schut y col., 1995; Williams y col., 2009). Se ha demostrado que en *S. alaskensis* RB2256 la glucosa es importada por un mecanismo secundario independiente de translocación de grupo (Schut y col., 1995).

También es importante mencionar que aunque la identificación de la familia a la que pertenece un transportador es relativamente fácil, determinar la especificidad de sustrato puede resultar complicado (Lee y col., 2008). Como consecuencia, el hecho de no detectar un transportador específico en la anotación funcional de un genoma, no supone que un

metabolito no pueda ser transportado al interior celular y, por tanto, se requiere de la demostración experimental para confirmarlo. En el caso de TFA, habría que analizar experimentalmente si la imposibilidad de utilizar glucosa se debe a la ausencia de transporte hacia el interior celular o a algún otro problema en los pasos sucesivos de la ruta catabólica, los cuales no se reflejan en el modelo metabólico.

### 2.1. Degradación de alcohol de polivinilo (PVA)

La biodegradación bacteriana del PVA ha sido ampliamente estudiada desde que en 1973 se mostrara que podía ser completamente degradado y utilizado como única fuente de carbono y energía por *Pseudomonas* O-3 (Suzuki y col., 1973). Desde entonces, se han definido distintos sistemas enzimáticos para la degradación de este polímero (Shimao, 2001).

Dentro del género *Sphingopyxis*, los genes encargados de la degradación de PVA se han caracterizado en *Sphingopyxis* sp. 113P3 (Klomklang y col., 2005; Hirota-Mamoto y col., 2006; Mamoto y col., 2008). El genoma de esta bacteria ha sido secuenciado y presenta, además de su cromosoma, un mega-plásmido de 243 kb (Ohtsubo y col., 2015). Los genes de degradación de PVA se encuentran formando un operón en dicho mega-plásmido (Hu y col., 2008) y codifican para la PVA deshidrogenasa (*pvdh*), para la hidrolasa de PVA oxidado (*oph*) y para el citocromo c (*cytC*). En 113P3, la degradación de PVA requiere de su incorporación al periplasma, donde la enzima PVA deshidrogenasa oxida dos grupos hidroxilos consecutivos del polímero, que pasan a formar una estructura  $\beta$ -dicetona. A continuación, se produce una rotura hidrolítica carbono-carbono entre los dos grupos carbonilos de la  $\beta$ -dicetona, la cual es catalizada por la hidrolasa de PVA oxidado (Kawai y Hu, 2009). La progresiva despolimerización del PVA por la acción de las enzimas anteriores puede llevar a la producción de ácido acético ( $\text{CH}_3\text{COOH}$ ) y ácidos grasos, que son transportados por la bacteria al interior de la célula para metabolizarlos a través de  $\beta$ -oxidación y el ciclo de Krebs (Kawai y Hu., 2009; Yang y col., 2014). La enzima PVA deshidrogenasa de 113P3 y la de otras bacterias degradadoras de PVA requieren del cofactor PQQ, así como de un citocromo c que actúa como aceptor de los electrones procedentes del PVA, los cuales son transportados a través del cofactor PQQ (Mamoto y col., 2008).

En TFA también se han identificado los genes de degradación de PVA, homólogos a los de 113P3 y formando también un operón (SGRAN\_2711, SGRAN\_2712 y SGRAN\_2713). Estos genes se han clasificado como genes exclusivos de TFA en relación con el género *Sphingopyxis*, si bien, como se ha visto previamente, la degradación de PVA no es específica de TFA dentro de este género. Esto se debe a que el análisis del pan-genoma del género *Sphingopyxis* se ha realizado sin incluir plásmidos y mega-plásmidos, que son los replicones

donde se suelen encontrar los genes de degradación de PVA (Hu y col., 2008) y, en concreto, es donde se encuentran en la estirpe 113P3. Teniendo en cuenta la dinámica de reorganizaciones génicas que suele afectar a los genomas de las esfingomonas y las evidencias de transferencia horizontal en el genoma de TFA, el operón de degradación de PVA podría tener un origen plasmídico, aunque en su contexto génico no aparecen genes relacionados con plásmidos, transposones o integrones. Es interesante señalar que en *S. granuli* ku-sg, la estirpe más parecida a TFA, no se han identificado los genes de degradación de PVA.

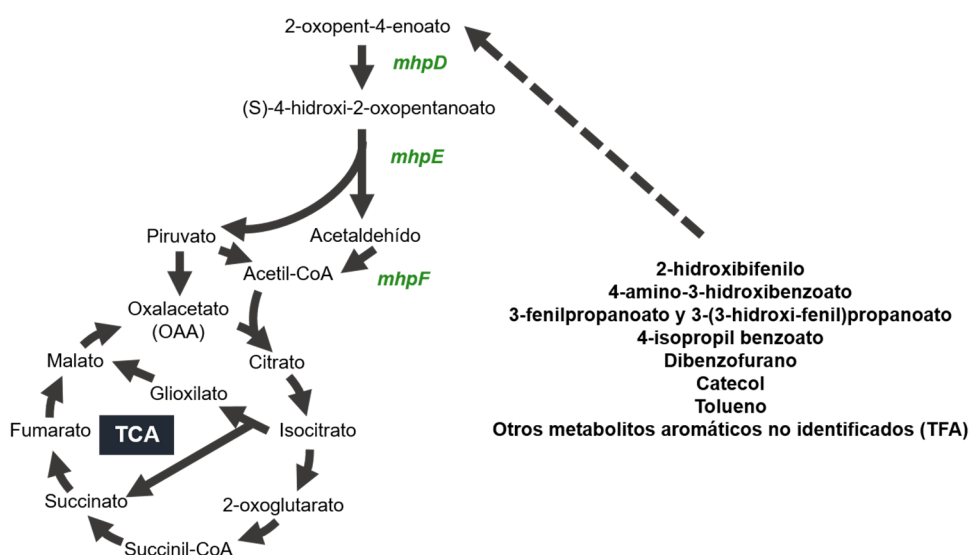
También cabe mencionar que, a pesar de la presencia de los genes para la degradación de PVA en el genoma de TFA, no se han identificado los genes encargados de la síntesis del cofactor PQQ, requerido por la enzima PVA deshidrogenasa. Aunque podría ocurrir que el cofactor de la PVA deshidrogenasa de TFA no fuera el PQQ, la ausencia de crecimiento utilizando PVA como única fuente de carbono y energía sugiere que sí podría depender de PQQ (datos no mostrados).

En la degradación de PVA se han descrito cultivos mixtos de bacterias que, en conjunto, son capaces de degradar el polímero si bien no pueden hacerlo de forma aislada (Sakazawa y col., 1981). Normalmente, los cultivos mixtos están formados por dos bacterias distintas. Una de ellas contiene los genes implicados en la degradación de PVA pero carece de la síntesis de PQQ, mientras que la otra suministra el PQQ a la primera pero no contiene las enzimas degradadoras. La degradación de PVA por cultivos mixtos se ha descrito, por ejemplo, entre *Pseudomonas* sp. VM15C y *Pseudomonas* sp. VM15A, basada en una relación de simbiosis, en la que una reacción química se lleva a cabo gracias a la intervención de ambas especies. En este caso, VM15C es la estirpe que degrada y metaboliza el PVA y VM15A quien secreta PQQ para que VM15C pueda catabolizar el PVA. VM15A recibe a cambio una serie de metabolitos indeterminados excretados por VM15C (Sakazawa y col., 1981; Shimao, 2001).

Debido a que en el genoma de TFA no están codificadas las enzimas requeridas para la síntesis de PQQ, es posible plantear que la degradación de PVA en esta bacteria ocurra a través de una relación simbiótica con otra que le proporcione el cofactor. Dentro del grupo de las esfingomonas ya se ha estudiado este tipo de degradación simbiótica. *Sphingomonas* sp. SA3 es capaz de degradar PVA en un medio en el que se suministra caldo de cultivo de *Sphingomonas* sp. SA2, que proporciona el cofactor PQQ (Kim y col., 2003).

## 2.2. Degradación de compuestos aromáticos

El genoma de TFA contiene dos agrupaciones parciales de genes *mhp*, los cuales están implicados en la degradación de distintos compuestos aromáticos en otras bacterias (Díaz y col., 2001). En concreto, las enzimas MhpD, MhpE y MhpF son las responsables de la ruta baja de la degradación de múltiples compuestos aromáticos cuyo catabolismo converge en 2-oxopent-4-enoato. Entre estos compuestos se encuentran el 3-fenilpropanoato y el 3-(3-hidroxi-fenil)propanoato, para los cuales se han identificado en el genoma de TFA posibles enzimas de la ruta alta de degradación. Sin embargo, TFA no mostró crecimiento en medio mínimo con estos metabolitos como únicas fuentes de carbono y energía. Este hecho, junto con que solo las enzimas de la ruta baja presentan una elevada similitud con las revisadas en la base de datos Swiss-Prot, sugieren la posibilidad de que TFA degrade por la ruta alta, también identificada, otros compuestos aromáticos similares a 3-fenilpropanoato y 3-(3-hidroxi-)fenilpropanoatos cuyo metabolismo genere 2-oxopent-4-enoato (Figura 65).



**Figura 65. Convergencia de la degradación de distintos compuestos aromáticos en el metabolito 2-oxopent-4-enotato.** La ruta alta de degradación de los compuestos representados es catalizada por distintas enzimas, mientras que la ruta baja es común a todos ellos y en ella están implicadas las enzimas codificadas en los genes *mhpD*, *mhpE* y *mhpF* que se indican en verde. Imagen adaptada de MetaCyc.

## 2.3. Síntesis de ectoína

En relación con la síntesis de metabolitos secundarios, en el modelo *ilG738* se ha reconstruido la ruta de biosíntesis del soluto compatible ectoína. Este compuesto es producido por los microorganismos para protegerse frente al estrés osmótico y se trata de uno de los solutos compatibles más encontrados en bacterias halófilas o halotolerantes (Pastor y col.,

2010). Además de la síntesis de ectoína, TFA puede producir hidroxiectoína. Esta molécula también actúa como soluto compatible pero, mientras que la ectoína actúa principalmente como osmoprotector, la hidroxiectoína también presenta un papel importante en la protección frente a estrés térmico (Pastor y col., 2010).

La síntesis de ectoína comienza con la fosforilación de L-aspartato y comparte las dos primeras reacciones enzimáticas con la ruta de biosíntesis de los aminoácidos de la familia del aspartato, que son la conversión de L-aspartato a L-aspartato-fosfato por la aspartato quinasa (Ask, SGRAN\_2349) y la síntesis de L-aspartato- $\beta$ -semialdehído desde L-aspartato-fosfato, catalizada por la enzima L-aspartato- $\beta$ -semialdehído deshidrogenasa (Asd, SGRAN\_0849). A partir de este intermediario comienza la ruta específica de biosíntesis de la ectoína. Inicialmente, el L-aspartato- $\beta$ -semialdehído se convierte a ácido L-aminobutírico por una transaminasa (EctB, SGRAN\_2346), el cual es posteriormente acetilado a ácido N- $\gamma$ -acetildiaminobutírico (NADA) por una acetil-transferasa (EctA, SGRAN\_2345). Por último, se forma la ectoína por la condensación cíclica de NADA catalizada por la ectoína sintasa (EctC, SGRAN\_2347). Para la síntesis de hidroxiectoína, la ectoína se hidroxila mediante una ectoína hidroxilasa (EctD, SGRAN\_2348).

En el genoma de TFA, todas las enzimas implicadas en la síntesis de ectoína e hidroxiectoína, salvo Asd, se encuentran codificadas en un operón (*ectABCDask*). Esta misma organización génica se ha encontrado en otras bacterias como la acidófila *Acidiphilium cryptum* (Moritz y col., 2015). Sin embargo, estos genes no siempre se organizan en un operón. Por ejemplo, en *Nitrosococcus oceani* se dividen en dos unidades transcripcionales, *ectABD* y *ectCask* (Pastor y col., 2010). A pesar de que *in vivo* no se ha probado la producción de ectoína e hidroxiectoína en TFA, la presencia de los genes para la biosíntesis de estos metabolitos sugiere que, en las condiciones adecuadas, podría producirlos.

La ectoína e hidroxiectoína no solo son importantes como protectores frente al estrés en los microorganismos, sino que pueden utilizarse en distintas aplicaciones biotecnológicas. Estos metabolitos tienen la capacidad de estabilizar las proteínas que son sometidas a condiciones de desnaturalización o altas temperaturas en los procesos industriales, presentan propiedades terapéuticas en ciertas enfermedades y protegen la piel de los efectos del daño celular causado por la radiación ultravioleta (revisado en Pastor y col., 2010). Como consecuencia, la producción a gran escala de ectoínas y la optimización del proceso ha suscitado un gran interés (Nagata y col., 2008; Guzmán y col., 2009; Fallet y col., 2010). La disponibilidad del modelo metabólico *i*LG738, permite analizar y optimizar la producción de ectoína en TFA e incluso diseñar mutantes con los que se consiga una mayor producción del metabolito.

### 2.4. Producción de gránulos de PHB y su efecto sobre el crecimiento

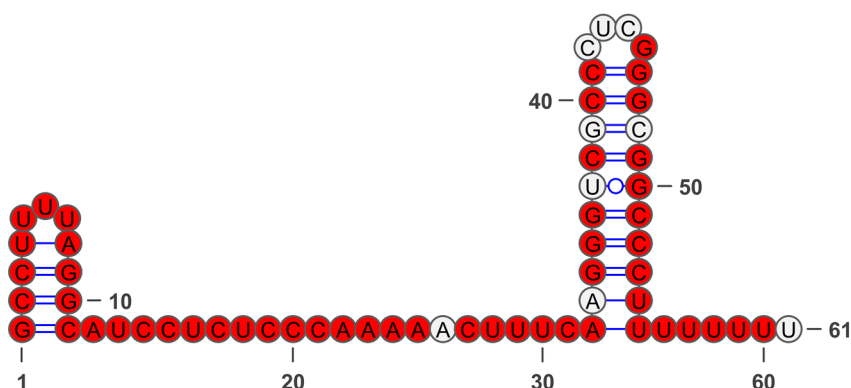
Otra de las capacidades metabólicas de TFA es la producción de gránulos de poli-hidroxibutirato (PHB), lo cual ha sido validado experimentalmente (Martín y col., 2011). Además, en el modelo *i*lG738 se ha reconstruido tanto la ruta de síntesis de PHB como la de su despolimerización. El modelo *i*lG738 ha sido capaz de simular el efecto de la acumulación de PHB en TFA creciendo tanto en aerobiosis como en anaerobiosis, con  $\beta$ -HB o ácido sebácico como fuentes de carbono (Figura 38 y Tabla 20). Como era de esperar, al incrementar la demanda de PHB el modelo redirige el flujo de carbono para la síntesis del polímero, en detrimento de la formación de biomasa, disminuyendo la velocidad de crecimiento. De hecho, en el análisis de flujos (FBA) se observa una disminución de flujo por las enzimas del ciclo de Krebs. Por tanto, según el modelo, la bacteria podría crecer mejor tanto aeróbica como anaeróbicamente si no acumulase PHB. Sin embargo, un mutante en *phaC*, incapaz de producir PHB crece más lentamente que la estirpe silvestre en cualquier condición. Este comportamiento opuesto entre lo predicho *in silico* y lo observado *in vivo* es debido a que en el segundo caso no se obliga a la bacteria a mantener una producción concreta de PHB y, en el caso en el que se vea afectado el crecimiento, la bacteria podría utilizar el polímero acumulado como fuente de carbono y energía, lo cual no se permite en el modelo metabólico. Además, se sabe que el PHB no sólo es útil como reserva de carbono y energía, sino que cumple multitud de funciones en la célula. El monómero del PHB (3-hidroxibutirato) ha mostrado la capacidad de proteger a las enzimas lipasa y lisozima de la desnaturalización por calor (Obruca y col., 2016). Por otra parte, el polímero protege frente a distintos estreses como alta temperatura, baja temperatura, congelación y estrés oxidativo (Pavez y col., 2009; Wu y col., 2011; Goh y col., 2014), por lo que el peor crecimiento del mutante puede deberse a efectos pleiotrópicos ajenos al mero cambio en el flujo metabólico de la fuente de carbono.

Por otra parte, tanto en el modelo como experimentalmente se observa que en anaerobiosis el crecimiento con ácido sebácico como fuente de carbono es más lento que con  $\beta$ -HB, mientras que ocurre lo contrario en aerobiosis. Atendiendo a la ruta catabólica de ambas fuentes de carbono, antes de incorporarse al ciclo de Krebs, el consumo de una molécula de  $\beta$ -HB (4 átomos de carbono) genera un NADH y un acetoacetyl-CoA, mientras que a partir de una molécula de ácido sebácico (10 átomos de carbono) se obtienen 3 NADH, 3 FADH<sub>2</sub>, 3 acetyl-CoA y un succinil-CoA tras los distintos ciclos de  $\beta$ -oxidación. Como consecuencia, la obtención de poder reductor es mayor en ácido sebácico que en  $\beta$ -HB, lo cual podría tener un mayor efecto sobre el crecimiento en anaerobiosis mientras que no resulte un problema en aerobiosis.



### 3. Implicación de un nuevo elemento en la regulación de los genes *thn*

El pequeño ARN SuhB, anotado en el genoma de *Sphingopyxis granuli* TFA por la herramienta Infernal y detectado experimentalmente mediante dRNA-seq y *Northern Blot*, es un elemento que, en otras bacterias donde se ha identificado, no se ha relacionado con los genes adyacentes sino que podría actuar en *trans* (Corbino y col., 2005). Se encuentra conservado en las bacterias de la clase  $\alpha$ -proteobacteria y, en esta tesis, se ha demostrado que dentro del género *Sphingopyxis* se conserva tanto su secuencia y estructura como su localización génica. Además, la secuencia conservada se extiende aguas arriba del inicio de la transcripción y llega hasta la región promotora e incluso a la zona donde podrían unirse proteínas reguladoras de su transcripción. Esta conservación sugiere que la expresión de *suhB* esté regulada por los mismos elementos dentro del género *Sphingopyxis* y que el pequeño ARN tenga dianas comunes en estas bacterias. Para confirmar esta última posibilidad se requieren más datos pero es importante destacar que la región 5' del pequeño ARN, donde suele localizarse la secuencia *seed* para la interacción con las dianas, es idéntica en todas las *Sphingopyxis* estudiadas (Figura 66). De hecho, la posible región de unión a Hfq también está muy conservada y donde aparecen más diferencias es únicamente en la región terminadora. Sin embargo, como se mostró en la sección de resultados, la estructura del terminador se conserva en todas ellas.



**Figura 66. Conservación del pequeño ARN SuhB en el género *Sphingopyxis*.** Se representa la secuencia de SuhB en *S. granuli* TFA. En rojo se marcan las bases que son idénticas entre las distintas estirpes analizadas del género *Sphingopyxis*.

Comparando las tres partes principales de un pequeño ARN codificado en *trans* (región *seed*, zona de interacción con Hfq y terminador), se observa una extraordinaria conservación de la secuencia de cada una de las partes. Puesto que la región *seed* controla el emparejamiento específico con el ARN diana, su conservación indica que SuhB controla la expresión de dianas comunes dentro del género *Sphingopyxis*, además del gen *thnR* específico de TFA. De forma similar, la gran conservación de la región central rica en As y Us,

## Discusión

la presumible zona de interacción con la proteína Hfq, probablemente refleje la gran conservación de esta proteína dentro del género *Sphingopyxis*. Por el contrario, un terminador independiente de Rho está constituido por una estructura secundaria seguida de una ristra de Us que aparecen en el extremo 3' del ARN transcrito, por lo que su secuencia puede cambiar sustancialmente siempre que se mantengan los criterios generales para ser considerado un terminador.

Aunque la región de estructura secundaria previa a la ristra de Us está algo menos conservada que las regiones precedentes, se aprecia también una conservación de secuencia muy considerable teniendo en cuenta que lo importante para un terminador es su estructura secundaria y no la secuencia concreta que forme dicha estructura. Probablemente, esta conservación de secuencia se deba a la estabilidad termodinámica de la estructura secundaria ya que ésta puede afectar al sitio concreto de terminación de la ARN polimerasa, influyendo por tanto en el número de Us que componen el extremo 3' del ARN. Aparentemente, este número de Us en el extremo 3' de ARNs que actúan en *trans* puede afectar a su capacidad de unir Hfq y, por tanto, a su funcionalidad, como se ha demostrado para SgrS y RyhB (Morita y col., 2017).

La función de SuhB no se ha estudiado en otras bacterias hasta el momento aunque sí se ha analizado su expresión en algunas  $\alpha$ -proteobacterias. En *Bradyrhizobium japonicum* se ha confirmado la expresión por *Northern Blot*, microarray y dRNA-seq de tres de los cinco parálogos identificados por homología pertenecientes a la familia RF00519 (BjrC2a, BjrC2b y BjrC2c). Con los datos de dRNA-seq se demostró que BjrC2a presenta un aumento notable de la expresión en los bacteroides que infectan a plantas de soja con respecto a células libres creciendo en fase exponencial (Madhugiri y col., 2012). En *Caulobacter crescentus*, uno de los cuatro parálogos de la familia RF00519 identificados, localizado entre los genes CC0741 y CC0743, se induce en fase estacionaria (Landt y col., 2008). A pesar de que estos pequeños ARN se han agrupado en la misma familia que el SuhB de TFA, tanto a nivel de secuencia como de estructura no son muy similares al SuhB definido en las bacterias del género *Sphingopyxis*, por lo que debemos suponer que regulan dianas distintas. Sin embargo, de todas las secuencias de SuhB almacenadas en la base de datos BSRD (un total de 178 a fecha de 12 de octubre de 2017), en 143 se conserva una pequeña secuencia de 5 nucleótidos (CUCCC), que comienza en el nucleótido 17 de la representación de la Figura 66, la cual podría ser clave para la funcionalidad de este tipo de pequeño ARN dentro de la clase  $\alpha$ -proteobacteria.

En la  $\alpha$ -proteobacteria *Sinorhizobium meliloti*, el pequeño ARN MmgR, de la misma familia que SuhB, se requiere para limitar la acumulación de gránulos de PHB en condiciones de

exceso de carbono. De hecho, el nombre asignado a este pequeño ARN se basa en que un mutante del gen *mmgR* da lugar a una estirpe que hace más gránulos (del inglés, *makes more granules*) (Lagares y col., 2016). La cantidad de este pequeño ARN es mayor en la fase estacionaria que en la fase exponencial, como ocurre con SuhB en *Caulobacter crescentus*. Se sabe, además, que la expresión de MmgR está regulada por la fuente de nitrógeno (Ceizel Borella y col., 2016).

Como se mencionó en la introducción, ya se han estudiado algunos pequeños ARNs reguladores que ejercen un papel fundamental en el control del fenómeno de represión catabólica en bacterias. En el caso de TFA, uno de los pequeños ARNs, SuhB, se expresa en medios de cultivo que permiten una elevada velocidad de crecimiento. Un mutante de TFA carente de SuhB muestra una desrepresión de los genes *thn* en condiciones de represión catabólica durante las primeras 15 horas de crecimiento, tiempo en el que la expresión se encuentra reprimida en la estirpe silvestre. La desrepresión observada en dicho mutante es menor que la que tiene lugar en un mutante carente de PHB, lo que indica el papel predominante de la síntesis de este polímero en la represión catabólica de los genes *thn*. El hecho de que en ninguno de los dos mutantes implicados hasta el momento en represión catabólica de los genes *thn* se alcancen los niveles de expresión de la estirpe silvestre en condiciones de inducción sugiere que algún otro proceso debe estar también implicado en la represión catabólica de estos genes.

El análisis *in silico* llevado a cabo por IntaRNA para detectar las posibles dianas de SuhB junto con los experimentos de retardo en gel, pusieron de manifiesto que una de las dianas de SuhB es el ARN mensajero de ThnR. En concreto, la interacción entre ambos ARNs ocurre en la región de unión del ribosoma del extremo 5' del ARNm, por lo que la traducción podría estar bloqueada. Esta interacción física se detectó en experimentos de EMSA y su previsto efecto negativo sobre la traducción de ThnR se confirmó tras la comprobación por Western Blot de que los niveles de proteína ThnR en una estirpe carente de SuhB son mayores que los de TFA silvestre en cualquier condición de crecimiento y especialmente en condiciones de represión catabólica de los genes *thn*. La misma tendencia en los niveles de ThnR se observó en el mutante de inserción en *phaC*, lo cual indica que los mecanismos de represión catabólica en este mutante afectan también, de forma directa o indirecta, a la síntesis de la proteína reguladora.

El crecimiento del mutante  $\Delta\text{suhB}::\text{kan}$  no se ve afectado en ninguna de las condiciones analizadas salvo cuando las células que provienen de un medio mínimo con  $\beta$ -HB 40 mM se pasan a un medio mínimo con tetralina. En esta condición, tanto la estirpe silvestre como el mutante presentan una fase de latencia para la adaptación al nuevo medio, pero esta fase es

## Discusión

mayor en TFA silvestre (de más de 10 horas) que en el mutante (de unas 6 horas) (Figura 51B). Este acortamiento del periodo de latencia se correlaciona con una mayor inducción de los genes *thn* en las primeras horas de crecimiento (Figura 52) y puede explicarse por la falta de SuhB en el mutante. En éste, la concentración del regulador sería más alta que en el silvestre, al no tener inhibida su traducción por SuhB de manera que en medio de inducción de los genes *thn* (con tetralina como única fuente de C y energía), ThnR podría empezar a activar la transcripción desde los promotores *thn* más rápidamente. Sin embargo, en TFA silvestre, al estar la traducción de ThnR reprimida por SuhB, la activación de la transcripción de los genes *thn* y, por tanto, el inicio del crecimiento utilizando tetralina como única fuente de C no se producirán hasta que la falta de expresión de *suhB* se refleje en una disminución de su concentración que permita una traducción efectiva de ThnR.

En base a lo descrito para los pequeños ARNs reguladores codificados en *trans* con una hibridación imperfecta con su diana, como es el caso de SuhB, la proteína Hfq podría estar implicada en la interacción y estabilización de la interacción entre SuhB y el extremo 5' del ARNm de *thnR*. Esta hipótesis se validó al comprobar que tanto el pequeño ARN como su diana co-inmunoprecipitan con Hfq-3xFLAG y que la expresión de la fusión traduccional *thnR::gfp* disminuye más en presencia de Hfq en la reconstrucción del sistema de forma heteróloga en *E. coli*. Como consecuencia, un mutante en el gen codificante para Hfq en TFA debería presentar un fenotipo de desrepresión de los genes *thn* parecido al observado en el mutante  $\Delta$ *suhB::kan*, puesto que la interacción entre SuhB y el mensajero de *thnR* sería menos eficaz, lo que permitiría la traducción, al menos en parte, del ARNm de *thnR* incluso en condiciones de represión catabólica.

Al igual que en *B. japonicum* y *C. crescentus*, *suhB* se expresa diferencialmente en TFA dependiendo de las condiciones de crecimiento, detectándose un alto nivel de SuhB en aerobiosis en presencia de una fuente de carbono preferencial y un menor nivel cuando utiliza tetralina como única fuente de carbono. Sin embargo, su expresión no se reprime solo cuando hay una fuente pobre de carbono y energía como la tetralina, sino que también se reprime mientras utiliza  $\beta$ -HB como fuente de carbono y energía cuando se ve obligada a respirar nitrato, que es una condición menos favorable desde un punto de vista energético que hace que su crecimiento se enlentezca mucho (tg=3,6 h en aerobiosis vs tg=12,5 h en anaerobiosis). Estos resultados sugieren que la expresión de *suhB* está controlada por el estado energético de la célula y no por la mera presencia de una fuente de carbono preferencial. En *Pseudomonas putida* ya se ha demostrado que el flujo de electrones a través de la cadena de transporte y el estado de oxidación-reducción (redox) de la célula podría ser una de las señales que la bacteria utilice para modular la represión catabólica de la ruta de

degradación de alcanos (Dinamarca y col., 2002). Del mismo modo, en TFA, el diferente estado redox entre aerobiosis y anaerobiosis o entre fuentes de carbono con diferente potencial energético podría ser una señal para la activación o represión de elementos implicados en la represión catabólica de los genes *thn*, como *SuhB*.

Desde un punto de vista fisiológico es razonable plantear que, en unas condiciones de crecimiento desfavorables, la célula inactive los mecanismos de represión catabólica para aumentar su potencial para obtener energía. Además, como se ha mencionado anteriormente, *SuhB* podría tener múltiples dianas en el genoma de TFA, por lo que no podemos descartar que su ausencia en condiciones de anaerobiosis sea apropiada para permitir la traducción de ARNm que codifican para proteínas necesarias en esta situación. Claro, que si se asume una situación de este tipo, la traducción de estas proteínas necesarias en anaerobiosis también se daría en aerobiosis en un medio con tetralina como única fuente de carbono y energía, por lo que debería existir otro nivel de regulación, probablemente a nivel transcripcional, que impida la expresión de estos genes en aerobiosis.

También se ha determinado que la expresión de *suhB* es menor en el mutante de inserción en *phaC* que en la estirpe silvestre tanto en medio mínimo con  $\beta$ -HB 40 mM como en medio mínimo con  $\beta$ -HB 40 mM y tetralina, si las células con las que se inoculan estos medios vienen adaptadas al crecimiento en tetralina. La falta de PHB en la célula y, por tanto, la menor disponibilidad de fuente de carbono, podría estar afectando a la expresión de *SuhB* con la misma tendencia que en un medio donde el rendimiento energético es menor (como en tetralina o anaerobiosis). En el caso de *Bacillus anthracis*, el PHB actúa como un reservorio endógeno de carbono y energía durante la fase de esporulación, de forma que un mutante incapaz de producir PHB reduce significativamente la eficiencia de la esporulación, debido a que se reduce el estado energético de la célula (Sadykov y col., 2017).

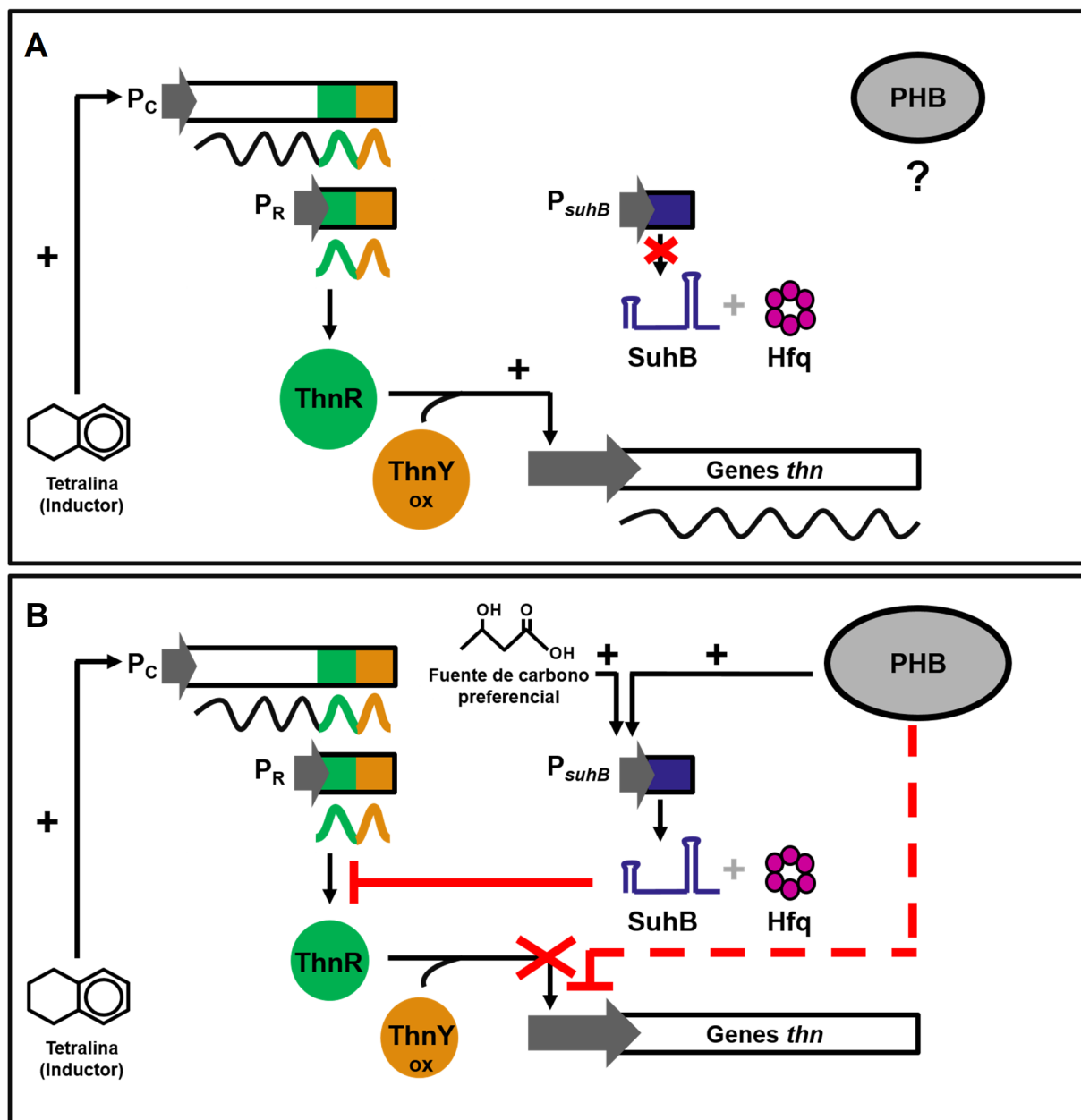
Además, partiendo de células crecidas en tetralina, el crecimiento del mutante de inserción en *phaC* se ve afectado, apreciándose una fase inicial de latencia hasta las 3 horas en medio mínimo con  $\beta$ -HB 40 mM y de unas 10 horas en medio mínimo con  $\beta$ -HB 40 mM y tetralina, mientras que en TFA silvestre no se observa fase de latencia. Este periodo de ausencia de crecimiento es incluso mayor si la única fuente de carbono y energía es la tetralina. Una explicación para este fenotipo podría ser que, sin la acumulación de gránulos de PHB, la bacteria fuera más sensible al efecto tóxico de la tetralina y que, en el periodo inicial de adaptación al crecimiento con tetralina (ver apartado 8.1.1 de materiales y métodos), las células se vieran más afectadas que el silvestre e incluso que una proporción muriera. Sin embargo, se requieren más datos experimentales, como el conteo de células viables en estas condiciones, para confirmar dicha hipótesis.

Finalmente, la información proporcionada en esta tesis doctoral permite actualizar el modelo de regulación de los genes *thn* incorporando nuevos elementos (Figura 67). Cuando la bacteria crece en un medio con una buena fuente de carbono como ciertos ácidos grasos o  $\beta$ -hidroxibutirato, la célula se encuentra en un estado energético favorable y los niveles del pequeño ARN SuhB son elevados. En estas condiciones, *thnR* se está transcribiendo constitutivamente a un nivel bajo desde su propio promotor  $P_R$  (López-Sánchez y col., 2009). Los elevados niveles de SuhB evitan la traducción de ThnR por lo que los niveles de esta proteína activadora son muy bajos. Además, en ausencia de tetralina, ThnR tampoco es activo, por lo que la transcripción desde los promotores *thn* es meramente basal. En presencia de estas fuentes de carbono preferenciales, aun en presencia de tetralina, que activaría la función de ThnR, los niveles de la proteína activadora serían tan bajos por la presencia de SuhB, que no se activaría la transcripción desde los promotores *thn*. Solo cuando la fuente de carbono preferente fuera limitante, bien por su consumo durante el crecimiento o por estar ausente, el estado energético deficitario de la célula evitaría la transcripción de *suhB*, lo que redundaría en una traducción de ThnR desde el transcrito producido desde su propio promotor  $P_R$ , incrementando su concentración. Si en estas condiciones de limitación energética existe el inductor específico tetralina, la función de ThnR se activará e iniciará la activación de la transcripción desde los promotores *thn*, incluido  $P_C$ , lo que lleva como consecuencia un incremento de la transcripción de *thnR*, que a su vez conduce a un aumento de la transcripción desde  $P_C$  (y desde los otros promotores *thn*), en un proceso de autoactivación, hasta alcanzar los niveles máximos de expresión desde los promotores *thn*.

Los niveles de expresión de *suhB* se ven incrementados por la capacidad de la célula de producir gránulos de PHB a través de un mecanismo cuya naturaleza molecular es desconocida hasta el momento, pero que puede tener que ver con un mejor estado energético de las células que puedan utilizar el PHB acumulado en sus gránulos. Sin embargo, la falta de producción de PHB debe, a su vez, afectar a la expresión de los promotores *thn* de manera independiente a SuhB, ya que el fenotipo de desrepresión catabólica del mutante de inserción en *phaC* es incluso más evidente que el del propio mutante *suhB*.

Por último, debe/n existir otro/s mecanismo/s de control que afecte/n a los niveles máximos de expresión de los promotores *thn* ya que los niveles de expresión de los mutantes *suhB* o *phaC* en presencia de una fuente de carbono preferencial y tetralina siempre son bastante menores que los que alcanzan cuando TFA crece con tetralina como única fuente de carbono y energía. Teniendo en cuenta que no ha sido posible obtener activación de la transcripción de los promotores *thn* *in vivo* en *E. coli* ni en un sistema de transcripción *in vitro*, aún proporcionando todos los elementos reguladores conocidos hasta la fecha, una

posibilidad es que en el modelo de regulación aún falte algún elemento necesario para la transcripción, como por ejemplo, un factor sigma alternativo, y que este nuevo elemento esté a su vez regulado en su síntesis o en su actividad por la disponibilidad de carbono o el estado energético de la célula.



**Figura 67. Modelo de regulación de la expresión de los genes *thn* en condiciones de inducción (A) y represión catabólica (B).** Las líneas rojas indican represión y la cruz roja ausencia de transcripción (A) o falta de activación (B). Las flechas negras y el símbolo "+" negro representan activación de la expresión y el símbolo "+" en gris representa acción conjunta. Las líneas discontinuas denotan falta de conocimiento del mecanismo molecular y la interrogación que no se ha determinado la producción de PHB en esas condiciones aunque se postula que su cantidad sea menor que en presencia de una fuente preferencial de carbono. Las líneas onduladas corresponden con ARNm señalándose en verde *thnR* y en amarillo *thnY*.





## **CONCLUSIONES**

---



1. El genoma de *Sphingopyxis granuli* estirpe TFA ha sufrido numerosos eventos de transferencia horizontal mediante los cuales ha adquirido diversas funcionalidades, como la capacidad para degradar el solvente orgánico tetralina, el potencial para degradar alcohol de polivinilo y la resistencia a metales pesados.
2. Los análisis filogenéticos a escala genómica han permitido reasignar la estirpe TFA a la especie *Sphingopyxis granuli*. A su vez muestran que el género *Sphingopyxis* es un grupo compacto del que *S. baekryungensis* DSM 16222 está excluido.
3. *Sphingopyxis granuli* estirpe TFA presenta características genómicas propias de organismos oligotrofos, tales como la presencia de un único operón ribosómico, un número reducido de ARNt, una alta representación de genes relacionados con el transporte y metabolismo de lípidos, y una baja presencia de profagos y secuencias CRISPRs.
4. La elevada similitud de *Sphingopyxis granuli* estirpe TFA con *Sphingopyxis granuli* estirpe ku-sg junto con la anotación de proteínas propias de bacterias endofíticas respaldan que TFA pudiera asociarse con plantas.
5. La reconstrucción a escala genómica del metabolismo de *Sphingopyxis granuli* TFA, representada en el modelo *ilG738*, ha puesto de manifiesto que las fuentes de carbono utilizadas por la bacteria son principalmente aminoácidos y ácidos grasos y que, además, presenta el potencial de degradar otros compuestos aromáticos distintos a la tetralina.
6. La robustez del modelo *ilG738* es lo suficientemente alta como para predecir el comportamiento de TFA y el mutante MPO209 en diferentes condiciones de cultivo, tales como el crecimiento anaeróbico, exclusivo de la especie *S. granuli*.
7. La conservación del pequeño ARN SuhB dentro del género *Sphingopyxis* sugiere que regule dianas comunes en estas bacterias.
8. La expresión de SuhB está regulada en *Sphingopyxis granuli* TFA de manera que sólo está presente en condiciones en las que las células presentan un buen estado energético, tales como la utilización de un metabolismo respiratorio aerobio con una fuente de carbono preferencial. Esta regulación está relacionada con la producción de gránulos de PHB.

## **Conclusiones**

9. La represión de los genes *thn* mediada por SuhB sucede a través de la interacción del pequeño ARN con la región de unión al ribosoma del ARNm del regulador transcripcional ThnR, lo cual bloquea la traducción. Esta unión se ve favorecida por la presencia de la chaperona Hfq.
10. Deben existir elementos reguladores diferentes a SuhB y a la producción de gránulo de PHB que controlen la expresión de los genes *thn* en condiciones de represión catabólica.

**ANEXOS**



## Anexo I. Oligonucleótidos para el ensamblaje del genoma de TFA

Oligonucleótido	Secuencia	Objetivo
129ic..189aleF	AAGGGTGGAAATGGACTCG	PCRs aleatorias
129ic..189aleR	TGTCCATCTTCGAAGGAACC	PCRs aleatorias
129icF_129ic168	GGGAGTCATGGTGTCTATGAAC	Prueba de unión 129ic-168
131..219aleF	TTCGGAACGAAAAAGTACGC	PCRs aleatorias
131..219aleR	GAACGTGAAGCACTCGAAATC	PCRs aleatorias
131F	CTTATTTTCGCACCGATTGC	Combinaciones para unir 243-131-152ic-251-189-158ic
137F	ATGTCGGCGTGGAGAAAC	Prueba de unión 137-155ic
137F_1	TGGAAGAAACCTCGGCG	Prueba de unión 137-155ic
137R	GACACCTCTCCCTTTCG	Prueba de unión 219-137
146..171aleF	CACCATCCTGCTGATGACC	PCRs aleatorias
146..171aleR	TCGAGCTGCTCATCACTGTC	PCRs aleatorias
146F_146176ic	GGGTTTCATCGACATCTTCAC	Prueba de unión 146-176ic
149aleF	GATTCCTTTCCCGACGAAAC	PCRs aleatorias
149aleR	GCCACATATCCTGACAGTCC	PCRs aleatorias
149icF_149ic220	GAATTCAGCAGCATTTTCC	Prueba de unión 149ic-220
150R	GTATCTCCTTCGTCGTCC	Prueba de unión 185-150
152icF	GTCTTCGTACCGTTTGCTC	Combinaciones para unir 243-131-152ic-251-189-158ic
152icR_168152ic	CGAATATCCGCATGATGGTC	Prueba de unión 168-152ic
155icF_155ic174ic	GGATCGGTGAAGAGCGAAG	Prueba de unión 155ic-174ic
155R	CGTATGGAGCGAGCAGAAC	Prueba de unión 137-155ic
155R_1	GCAGACGACGAAGACAGC	Prueba de unión 137-155ic
156aleF	GAATCACCCATGGCACTCTC	PCRs aleatorias
156aleR	AGCTTTTCGACGTTTATTTTCG	PCRs aleatorias
158_5_FS	AAGGTCAAGGCGTTCAGG	Sonda del extremo 158_5'
158_5_RS	ACCACCGCCTATCACACG	Sonda del extremo 229_5'
158icR	GAATTGAGCGTCAGGCTTTC	Combinaciones para unir 243-131-152ic-251-189-158ic
158R	ATAGTTCAGCGGCACCAC	Prueba de unión 158ic-179
158R_SecCos	TCGTATTTGGCGCTGACC	Secuenciación cósmido/s 158_5'
158R_SecCos_1	TGACGGTGTGATAGAGCGAG	Secuenciación cósmido/s 158_5'
162F	CGAACGGGAGATGTGAGG	Prueba de unión 162-217
162R	GTGTTTCATCAACTGGTTCGG	Prueba de unión 244-162
163icF_163ic171	AGGGGAGAGGATAGCGTAGC	Prueba de unión 163ic-171
163R	CGTGAAGGTGATGGCGAAG	Prueba de unión 166-163ic
166F	TCATCTGGTTGCTGTTCGC	Prueba de unión 166-163ic
166icF_166ic191	GTGGCTCGACGAACAGAAC	Prueba de unión 166ic-191
168F_168152ic	TTCGATGTGGGAAATGC	Prueba de unión 168-152ic
168R_129ic168	CAGCAGGATATCAAGGACCAG	Prueba de unión 129ic-168
170F_170223ic	CTGCGAAGGGCTTTCCTAC	Prueba de unión 170-223ic
170icR_179170ic	GGCTTTTCGACCGTCTTC	Prueba de unión 179-170ic
170R_New	CGCAGACGCTGTTTCAGAC	Prueba de unión 223ic-170
171R_163ic171	CGAGATGGAGGATGTGGAG	Prueba de unión 163ic-171
174icR_155ic174ic	GATTCAGCGTCAGCTTGC	Prueba de unión 155ic-174ic
174R_235174	CATCTACGTTTTCGGGATG	Prueba de unión 235-174
176icR_146176ic	GCATCCTCGTCAACATCTTC	Prueba de unión 146-176ic

## Anexos

Oligonucleótido	Secuencia	Objetivo
176R_191176	CGTCCTTCGAGCGTGTAGG	Prueba de unión 191-176
179F	TCGGGCTCAACTGGAACG	Prueba de unión 158ic-179
179F_179170ic	CGCGCAACATTATTTC AAG	Prueba de unión 179-170ic
185_5	GTCGAAGAAGCGGCGGAGCAC	Prueba de unión 241ic-185
185150aleF	TTGTACCGCTTCTGGAGAGTG	PCRs aleatorias
185150aleR	AAAGCCGAACAGGATCTGG	PCRs aleatorias
185F	ATCCCGACTTTTGCTTTG	Prueba de unión 185-150
189R	TCGTGCAAATCATCTTCGAG	Combinaciones para unir 243-131-152ic-251-189-158ic
191F_191176	GACGATGGAGGTGCTGAG	Prueba de unión 191-176
191R_166ic191	CAAGTCCGGAATGACGAAG	Prueba de unión 166ic-191
217R	GTCGTGGAAGCGGCAGGC	Prueba de unión 162-217
219F	TGAATCCCTTTACGGTCTG	Prueba de unión 219-137
220aleR	TGCGAGGCTTACTACGAACC	PCRs aleatorias
220R_149ic220	TCCATAAGCTGGCGTAATCC	Prueba de unión 149ic-220
223_3_COSM	GTCGCCGAAATCATAGGTGTAG	Secuenciación cósmido/s 223_3'
223_3_COSM2	CAGGATCAGCAGATCGAAGC	Secuenciación cósmido/s 223_3'
223_3_COSM3	ATGTGCAAGGCTGAGCAAC	Secuenciación cósmido/s 223_3'
223_3_COSM4	GACGATAACCTCGGGTAAAGC	Secuenciación cósmido/s 223_3'
223_3_FS	CATCTGACGCAATCAGTCG	Sonda del extremo 223_3'
223_3_RS	TCCATGGTGCAATGGTATG	Sonda del extremo 223_3'
223aleF	GCCAAAGCCAAGTTTTTACG	PCRs aleatorias
223icR_170223ic	TGGATGACCTGTTCGTTGG	Prueba de unión 170-223ic
229..243aleF	CAAGCAGATCGGCAAATTCG	PCRs aleatorias
229..243aleR	TTATATCCATGGCGCACGTAG	PCRs aleatorias
229-248_FS	CTGACGCGTCTCTATCCTGTC	Sonda del extremo 229_5'
229-248_RS	AACTCGAGAAGGCTCTCAACC	Sonda del extremo 229_5'
229F_229242	GTTGGATCATCGGCCTTC	Prueba de unión 229-242
230icR	AGTTCAGGGTGACGAAAGG	Prueba de unión 237-230ic
232icF	CCTCGATATCCTGCTCTGG	Prueba de unión 232ic-238
232icR	AGCTATCAGCCGATGAACG	Prueba de unión 245ic-232ic
233F	CGCTTCCGCTCTTTTCTTAC	Prueba de unión 233-235
233icF_New	CACGGTGTGAGATGTGC	Prueba de unión 223ic-170
235F_235174	GAACCGATGTAGCGTGATTC	Prueba de unión 235-174
235R	GAGGATTCGCTGACACTGGC	Prueba de unión 233-235
236F	TAACGACACCACGAAGAACC	Prueba de unión 236-237
236R	GTTGCCGAGTCCGATGAAGC	Prueba de unión 251-236
237F	TCGACCTGCTGCTTTAGACTG	Prueba de unión 237-230ic
237R	GCCACCTTGTCAGTACATCC	Prueba de unión 236-237
238F	GCGCTCGACAAGAAATACAAG	Prueba de unión 238-240
238R	GCTATCAGGGATATGCGACAC	Prueba de unión 232ic-238
240R	GTAGAGGATGACGAGCTGCAC	Prueba de unión 238-240
241_3	CAGCAGGTTGGATCGGCGACGC	Prueba de unión 241ic-185
241aleF	AGGTCTGGGTTTGCAATGAC	PCRs aleatorias
241aleR	AACGATATTCGGGTGTCACG	PCRs aleatorias
242F	GACGGCTATGCTGCAATATG	Prueba de unión 242-243
242R_229242	CGGAACGACTGACTGACG	Prueba de unión 229-242



Oligonucleótido	Secuencia	Objetivo
<b>243F</b>	CAAGCAGATCGGCAAATTC	Combinaciones para unir 243-131-152ic-251-189-158ic
<b>243R</b>	TCGGAAGCAGGAAAAGACAC	Prueba de unión 242-243
<b>244..217aleF</b>	AGTAGCGGCTTACGCAGTTC	PCRs aleatorias
<b>244..217aleR</b>	TGATGGGCTGAGATTTCCTC	PCRs aleatorias
<b>244F</b>	ATCATCCCCTTCGTCTCGTC	Prueba de unión 244-162
<b>245_3</b>	GTGGCGAATGACGTGAAAC	Prueba de unión 245-248
<b>245F</b>	GCGTTTTTGCGATTGAGAGC	Prueba de unión 245-248
<b>245icF</b>	CGGGATGAGGTTGAGTGTC	Prueba de unión 245ic-232ic
<b>248_5</b>	CAAGTATGCGGATCATCTGC	Prueba de unión 245-248
<b>248ic..240aleF</b>	TGGGGAATGAAGATGACTCC	PCRs aleatorias
<b>248ic..240aleR</b>	TTCACGATCGTAAGCGTGTC	PCRs aleatorias
<b>248R</b>	CGTCGCCTATGTCTATGCTC	Prueba de unión 245-248
<b>251..230icaleF</b>	GATCAGCCACGGTCATCAAC	PCRs aleatorias
<b>251..230icaleR</b>	CGTAAAGGTCAAAC TGCCAAC	PCRs aleatorias
<b>251F</b>	AGGCGAAGGCGATGGAGATG	Prueba de unión 251-236
<b>251R</b>	GCCAGTCGACCTTG TAGCTG	Combinaciones para unir 243-131-152ic-251-189-158ic

**Anexo II. Listado de genomas extraídos de la base de datos GenBank y utilizados en los análisis de genómica comparada**

Genoma	Completo/ Parcial	Nº de fragmentos	Número de acceso
<i>Blastomonas</i> sp. AAP53	Parcial	28	NZ_ANFZ01000001-NZ_ANFZ01000028
<i>Erythrobacter litoralis</i> HTCC2594	Completo	1	NC_007722.1
<i>Novosphingobium aromaticivorans</i> DSM 12444	Completo	1	NC_007794.1
<i>Novosphingobium</i> sp. PP1Y	Completo	1	NC_015580.1
<i>Oligotropha carboxidovorans</i> OM4	Completo	1	NC_017538.1
<i>Oligotropha carboxidovorans</i> OM5	Completo	1	NC_015684.1
<i>Sphingobium chlorophenolicum</i> L-1	Completo	2	NC_015593.1, NC_015594.1
<i>Sphingobium japonicum</i> UT26S	Completo	2	NC_014006.1, NC_014013.1
<i>Sphingobium</i> sp. SYK-6	Completo	1	NC_015976.1
<i>Sphingomonas sanxanigenens</i> NX02	Completo	1	NZ_CP006644.1
<i>Sphingomonas</i> sp. MM-1	Completo	1	NC_020561.1
<i>Sphingomonas wittichii</i> RW1	Completo	1	NC_009511.1
<i>Sphingopyxis alaskensis</i> RB2256	Completo	1	NC_008048.1
<i>Sphingopyxis baekryungensis</i> DSM 16222	Parcial	10	NZ_ATUR01000001-NZ_ATUR01000010
<i>Sphingopyxis fribergensis</i> Kp5.2	Completo	1	NZ_CP009122.1
<i>Sphingopyxis granuli</i> ku-sg	Parcial	1	NZ_CP019153.1
<i>Sphingopyxis granuli</i> TFA	Completo	1	CP012199 (Esta Tesis)
<i>Sphingopyxis</i> sp. C1	Parcial	2	BBRO01000001-BBRO01000002
<i>Sphingopyxis</i> sp. LC363	Parcial	73	NZ_JNFC01000001-NZ_JNFC01000073
<i>Sphingopyxis</i> sp. LC81	Parcial	48	NZ_JNFD01000001-NZ_JNFD01000048
<i>Sphingopyxis</i> sp. MC1	Parcial	24	NZ_AOUN01000001-NZ_AOUN01000024
<i>Sphingopyxis</i> sp. MWB1	Parcial	5	NZ_JQFJ01000001-NZ_JQFJ01000005
<i>Zymomonas mobilis subsp. mobilis</i> ATCC 10988	Completo	1	NC_017262.1

### Anexo III. Código fuente en lenguaje Perl del programa utilizado para extraer 150 pb aguas arriba y 150 pb aguas abajo del inicio de la región codificante de los genes de TFA

```
#!/usr/bin/perl
use strict;
use warnings;
my $genoma;
my $s;
my $rs;
#abrir los ficheros de entrada con la anotación y la secuencia del genoma
open annot1, "./annot_TFA.txt";
open genom1, "./TFA_genome_20-11-13.fna";
#abrir el fichero de salida donde se guarda el resultado
open out1, ">5'utr-orf.fasta";
#almacenar la secuencia del genoma en la variable $genoma
while (<genom1>) {
    chomp;
    if ($_ !~ "^\>") {
        $genoma .= $_;
    }
}
#eliminar espacios y saltos de línea en la secuencia del genoma
$genoma =~ 's/\s//g';
$genoma =~ 's/\n//g';
#recorrer el fichero de la anotación y extraer las secuencias deseadas
while (<annot1>) {
    chomp;
    my ($ctg, $id, $type, $p1, $p2, $st, $gn, $des, $go, $ip, $key, $path, $nt, $aa,
    $r) = split /\t/;
    if ($st eq "+") {
        $s = substr $genoma, $p1-151, 301;
        $s =~ tr/T/U/;
        print out1 ">$id\_150-5'UTR+150-cod\n$s\n";
    }
    if ($st eq "-") {
        #para el primer gen
        if ($id eq "Gene_1") {
            $s = substr $genoma, $p1-151, 301;
            $s =~ tr/ATCG/UAGC/;
            $rs = reverse $s;
            print out1 ">$id\_150-5'UTR+150-cod\n$rs\n";
        }
        #para el último gen (no hay más secuencia aguas arriba del inicio)
        } elsif ($id eq "Gene_4298") {
            $s = substr $genoma, $p1-151, 151;
            print length($genoma);
            $s =~ tr/ATCG/UAGC/;
            $rs = reverse $s;
            print out1 ">$id\_150-5'UTR+150-cod\n$rs\n";
        } else {
            $s = substr $genoma, $p1-151, 301;
            $s =~ tr/ATCG/UAGC/;
            $rs = reverse $s;
            print out1 ">$id\_150-5'UTR+150-cod\n$rs\n";
        }
    }
}
#cerrar los distintos ficheros
close annot1;
close genom1;
close out1;
```

**Anexo IV. Total de islas genómicas predichas por IslandViewer 3**IslandPath-DIMOB

Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
1	549355	560431	11076	0512	Citrate lyase
1	549355	560431	11076	0513	Aldehyde dehydrogenase
1	549355	560431	11076	0514	Uncharacterized protein
1	549355	560431	11076	0515	Putative epoxide hydrolase
1	549355	560431	11076	0516	TonB-dependent receptor
1	549355	560431	11076	0517	IS21 family transposase
1	549355	560431	11076	0518	IstB domain protein ATP-binding protein
1	549355	560431	11076	0519	Uncharacterized protein
1	549355	560431	11076	0520	Regulatory protein MarR
1	549355	560431	11076	0521	Virulence-associated protein E
2	1691609	1760539	68930	1557	Integrase
2	1691609	1760539	68930	1558	Uncharacterized protein
2	1691609	1760539	68930	1559	Nuclease
2	1691609	1760539	68930	1560	Lytic transglycosylase
2	1691609	1760539	68930	1561	Uncharacterized protein
2	1691609	1760539	68930	1562	Phage transcriptional regulator, AlpA
2	1691609	1760539	68930	1563	Uncharacterized protein
2	1691609	1760539	68930	1566	Long-chain fatty acid--CoA ligase
2	1691609	1760539	68930	1567	Transposase
2	1691609	1760539	68930	1568	IS511, transposase OrfB
2	1691609	1760539	68930	1569	Thioesterase
2	1691609	1760539	68930	1570	Peroxidase
2	1691609	1760539	68930	1571	Marr family regulator
2	1691609	1760539	68930	1572	Ubiquinone biosynthesis protein
2	1691609	1760539	68930	1573	NUDIX hydrolase
2	1691609	1760539	68930	1574	NADPH:quinone oxidoreductase
2	1691609	1760539	68930	1575	Extradiol ring-cleavage dioxygenase
2	1691609	1760539	68930	1576	Protocatechuate 3,4-dioxygenase
2	1691609	1760539	68930	1577	Ring-hydroxylating dioxygenase large subunit
2	1691609	1760539	68930	1578	3-phenylpropionate dioxygenase subunit beta
2	1691609	1760539	68930	1579	4-hydroxy-2-keto-pentanoic acid aldolase
2	1691609	1760539	68930	1580	Acetaldehyde dehydrogenase 2/3
2	1691609	1760539	68930	1581	2-keto-4-pentenoate hydratase
2	1691609	1760539	68930	1582	Pca regulon regulatory protein
2	1691609	1760539	68930	1583	2,3-dihydroxy-2,3-dihydrophenylpropionate dehydrogenase
2	1691609	1760539	68930	1584	2-hydroxy-6-oxo-2,4-heptadienoate hydrolase
2	1691609	1760539	68930	1585	AMP-dependent synthetase and ligase
2	1691609	1760539	68930	1586	TonB-dependent receptor
2	1691609	1760539	68930	1587	Ferredoxin--NAD(P)(+) reductase fdr
2	1691609	1760539	68930	1588	putative insertion sequence transposase protein, IS3/IS911
2	1691609	1760539	68930	1589	Transposase
2	1691609	1760539	68930	1590	Mobile element protein
2	1691609	1760539	68930	1591	Integrase catalytic region
2	1691609	1760539	68930	1592	Transposase
2	1691609	1760539	68930	1593	TonB-dependent receptor
2	1691609	1760539	68930	1594	Acyl-CoA dehydrogenase family protein
2	1691609	1760539	68930	1595	Acyl-CoA dehydrogenase family protein
2	1691609	1760539	68930	1596	2Fe-2S ferredoxin
2	1691609	1760539	68930	1597	Cytochrome P450

Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
2	1691609	1760539	68930	1598	Regulatory protein GntR, HTH
2	1691609	1760539	68930	1599	Oxidoreductase, 2-nitropropane dioxygenase family protein
2	1691609	1760539	68930	1600	Uncharacterized protein
2	1691609	1760539	68930	1601	Putative nonspecific lipid-transfer protein
2	1691609	1760539	68930	1602	Putative nucleic-acid-binding protein containing a Zn-ribbon
2	1691609	1760539	68930	1603	Putative acyl-CoA dehydrogenase family protein
2	1691609	1760539	68930	1604	Putative acyl-CoA dehydrogenase
2	1691609	1760539	68930	1605	3-oxocholest-4-en-26-oate--CoA ligase
2	1691609	1760539	68930	1606	Cyclohexanone monooxygenase
2	1691609	1760539	68930	1607	Hydrolase, putative
2	1691609	1760539	68930	1608	TonB-dependent receptor
2	1691609	1760539	68930	1609	RND transporter
2	1691609	1760539	68930	1610	DSBA oxidoreductase
2	1691609	1760539	68930	1611	Multidrug ABC transporter permease
2	1691609	1760539	68930	1612	ABC-type transport system periplasmic component
2	1691609	1760539	68930	1613	Uncharacterized protein
3	1877102	1886676	9574	1712	Alpha/beta hydrolase fold protein
3	1877102	1886676	9574	1713	Uncharacterized protein
3	1877102	1886676	9574	1715	Integrase
3	1877102	1886676	9574	1716	Lytic transglycosylase
3	1877102	1886676	9574	1717	Uncharacterized protein
3	1877102	1886676	9574	1718	Putative methylase/helicase
3	1877102	1886676	9574	1719	Phage integrase
3	1877102	1886676	9574	1720	Phage integrase
4	2001817	2012061	10244	1834	Uncharacterized protein
4	2001817	2012061	10244	1835	Uncharacterized protein
4	2001817	2012061	10244	1836	HNH endonuclease
4	2001817	2012061	10244	1837	Terminase
4	2001817	2012061	10244	1838	Integrase
4	2001817	2012061	10244	1839	Resolvase domain
4	2001817	2012061	10244	1840	Putative transcriptional regulator
4	2001817	2012061	10244	1841	hypothetical protein
5	2434819	2449730	14911	2240	TonB-dependent receptor
5	2434819	2449730	14911	2241	Anti-FecI sigma factor, FecR
5	2434819	2449730	14911	2242	RNA polymerase, sigma-24 subunit, ECF subfamily
5	2434819	2449730	14911	2243	Conserved Putative secreted protein
5	2434819	2449730	14911	2244	LuxR-family transcriptional regulator
5	2434819	2449730	14911	2245	ATPase
5	2434819	2449730	14911	2246	Uncharacterized protein
5	2434819	2449730	14911	2247	hypothetical protein
5	2434819	2449730	14911	2248	Transcriptional regulator
5	2434819	2449730	14911	2249	Hemimethylated DNA-binding region
6	3066560	3103335	36775	2824	Uncharacterized protein
6	3066560	3103335	36775	2825	Uncharacterized protein
6	3066560	3103335	36775	2826	Uncharacterized protein
6	3066560	3103335	36775	2827	Uncharacterized protein
6	3066560	3103335	36775	2828	Uncharacterized protein
6	3066560	3103335	36775	2829	Cation transporter
6	3066560	3103335	36775	2830	Cation efflux system protein CusB
6	3066560	3103335	36775	2831	Heavy metal RND transporter
6	3066560	3103335	36775	2832	Methyltransferase type 12
6	3066560	3103335	36775	2834	Cation transporter

## Anexos

Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
6	3066560	3103335	36775	2835	Putative transcriptional regulator
6	3066560	3103335	36775	2836	Uncharacterized protein
6	3066560	3103335	36775	2837	Membrane protein
6	3066560	3103335	36775	2838	Conjugal transfer pilus assembly protein TraL
6	3066560	3103335	36775	2839	Conjugal transfer pilus assembly protein TraE
6	3066560	3103335	36775	2840	Conjugal transfer protein TraK
6	3066560	3103335	36775	2841	Conjugal transfer protein TraB
6	3066560	3103335	36775	2842	Thiol:disulfide interchange protein DsbC
6	3066560	3103335	36775	2843	Conjugal transfer pilus assembly protein TraV
6	3066560	3103335	36775	2844	Conjugal transfer protein TraC
6	3066560	3103335	36775	2845	Conjugative transfer protein
6	3066560	3103335	36775	2846	Uncharacterized protein
6	3066560	3103335	36775	2847	Conjugal transfer pilin signal peptidase TrbI
6	3066560	3103335	36775	2848	Conjugal transfer pilus assembly protein TraW
6	3066560	3103335	36775	2849	Conjugal transfer protein
6	3066560	3103335	36775	2850	Conjugal transfer pilus assembly protein TrbC
6	3066560	3103335	36775	2851	Conjugal transfer mating pair stabilization protein TraN
6	3066560	3103335	36775	2852	Conjugal transfer pilus assembly protein TraF
6	3066560	3103335	36775	2853	Conjugal transfer protein TraH
6	3066560	3103335	36775	2854	Conjugal transfer protein TraG
6	3066560	3103335	36775	2855	Superfamily I DNA/RNA helicase
6	3066560	3103335	36775	2856	Uncharacterized protein
6	3066560	3103335	36775	2857	Alkaline phosphatase
6	3066560	3103335	36775	2858	Helicase
6	3066560	3103335	36775	2859	Putative transposase
6	3066560	3103335	36775	2860	Isocitrate lyase
7	3109158	3161042	51884	2866	DNA methylase
7	3109158	3161042	51884	2867	Transposase IS3/IS911 family protein
7	3109158	3161042	51884	2868	Trnasposase
7	3109158	3161042	51884	2869	DNA methylase
7	3109158	3161042	51884	2870	ATPase
7	3109158	3161042	51884	2871	tRNA delta(2)-isopentenylpyrophosphate transferase
7	3109158	3161042	51884	2872	Transcriptional regulator
7	3109158	3161042	51884	2873	Putative phage repressor
7	3109158	3161042	51884	2874	Uncharacterized protein
7	3109158	3161042	51884	2875	Uncharacterized protein
7	3109158	3161042	51884	2876	Multifunctional conjugation protein TraI
7	3109158	3161042	51884	2877	Type IV secretory pathway VirD4 components-like protein
7	3109158	3161042	51884	2878	Uncharacterized protein
7	3109158	3161042	51884	2879	Uncharacterized protein
7	3109158	3161042	51884	2880	Uncharacterized protein
7	3109158	3161042	51884	2881	Uncharacterized protein
7	3109158	3161042	51884	2882	Single-stranded DNA-binding protein
7	3109158	3161042	51884	2883	Uncharacterized protein
7	3109158	3161042	51884	2884	Uncharacterized protein
7	3109158	3161042	51884	2885	Chromosome segregation ATPase
7	3109158	3161042	51884	2886	Uncharacterized protein
7	3109158	3161042	51884	2887	Antirestriction protein ArdC
7	3109158	3161042	51884	2888	Antirestriction protein
7	3109158	3161042	51884	2889	Phage integrase family protein
7	3109158	3161042	51884	2890	Uncharacterized protein
7	3109158	3161042	51884	2891	Putative phage type integrase/recombinase
7	3109158	3161042	51884	2892	Antirestriction protein

Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
7	3109158	3161042	51884	2893	ParB-like partition protein
7	3109158	3161042	51884	2894	Uncharacterized protein
7	3109158	3161042	51884	2895	DEAD/DEAH box helicase
7	3109158	3161042	51884	2896	hypothetical protein
7	3109158	3161042	51884	2897	Uncharacterized protein
7	3109158	3161042	51884	2898	RES domain protein
7	3109158	3161042	51884	2899	Uncharacterized protein
7	3109158	3161042	51884	2900	Putative methylase/helicase
7	3109158	3161042	51884	2901	Phage integrase
7	3109158	3161042	51884	2902	Phage integrase
7	3109158	3161042	51884	2903	Phage integrase
7	3109158	3161042	51884	2904	Integrase/recombinase
7	3109158	3161042	51884	2905	Integrase/recombinase
7	3109158	3161042	51884	2906	Putative integrase/recombinase y4gC
7	3109158	3161042	51884	2907	Putative methylase/helicase
7	3109158	3161042	51884	2908	Virulence-associated protein E
7	3109158	3161042	51884	2909	Lytic transglycosylase

## SIGI-HMM

Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
1	1896	9255	7359	0002	Uncharacterized protein
1	1896	9255	7359	0003	Uncharacterized protein
1	1896	9255	7359	0004	Uncharacterized protein
1	1896	9255	7359	0005	Asparagine synthase
1	1896	9255	7359	0006	Dipeptidyl aminopeptidases/acylaminoacyl-peptidases-like protein
2	42649	48748	6099	0037	Uncharacterized protein
2	42649	48748	6099	0038	Uncharacterized protein
2	42649	48748	6099	0039	dCTP deaminase
2	42649	48748	6099	0040	Uncharacterized protein
2	42649	48748	6099	0041	Phospholipase/Carboxylesterase
2	42649	48748	6099	0042	Cation diffusion facilitator family transporter
2	42649	48748	6099	0043	Uncharacterized protein
3	129619	134249	4630	0118	Uncharacterized protein
3	129619	134249	4630	0119	Uncharacterized protein
3	129619	134249	4630	0120	TonB-dependent receptor
3	129619	134249	4630	0121	Cytochrome c family protein
3	129619	134249	4630	0122	Uncharacterized protein
4	159051	176784	17733	0146	Histidine utilization repressor
4	159051	176784	17733	0147	Phage integrase family protein
4	159051	176784	17733	0149	Uncharacterized protein
4	159051	176784	17733	0150	hypothetical protein
4	159051	176784	17733	0151	Uncharacterized protein
4	159051	176784	17733	0152	Uncharacterized protein
4	159051	176784	17733	0153	Heat shock protein Hsp20
4	159051	176784	17733	0154	ETC complex I subunit region
4	159051	176784	17733	0155	Uncharacterized protein
4	159051	176784	17733	0156	Hydratase/decarboxylase family protein
4	159051	176784	17733	0157	Carbamoyl-phosphate synthase, large subunit family protein
4	159051	176784	17733	0158	Uncharacterized protein
4	159051	176784	17733	0159	4-hydroxy-tetrahydronicotinate synthase
4	159051	176784	17733	0160	Uncharacterized protein
4	159051	176784	17733	0161	Uncharacterized protein
4	159051	176784	17733	0162	Potassium-transporting ATPase A chain
4	159051	176784	17733	0163	Uncharacterized protein
5	178969	189781	10812	0165	Potassium-transporting ATPase C chain
5	178969	189781	10812	0166	Putative sensor protein KdpD
5	178969	189781	10812	0167	KDP operon transcriptional regulatory protein KdpE
5	178969	189781	10812	0168	Nitrogen regulatory protein PtsN
5	178969	189781	10812	0169	Molecular chaperone Hsp20
5	178969	189781	10812	0170	Transcriptional regulator, LysR family
5	178969	189781	10812	0171	Uncharacterized protein
5	178969	189781	10812	0172	Conjugal transfer protein TrbI
5	178969	189781	10812	0173	Conjugal transfer protein TrbG
5	178969	189781	10812	0174	Conjugal transfer protein TrbF
5	178969	189781	10812	0175	Conjugal transfer protein TrbL
5	178969	189781	10812	0176	Conjugal transfer protein TrbK
6	219786	226418	6632	0207	Single-stranded DNA endonuclease
6	219786	226418	6632	0208	Antirepressor
6	219786	226418	6632	0209	Antirestriction protein



Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
6	219786	226418	6632	0210	Phage integrase family protein
6	219786	226418	6632	0211	Uncharacterized protein
6	219786	226418	6632	0212	Putative phage type integrase/recombinase
7	230013	235133	5120	0216	Uncharacterized protein
7	230013	235133	5120	0217	DEAD/DEAH box helicase
7	230013	235133	5120	0218	Uncharacterized protein
7	230013	235133	5120	0219	Uncharacterized protein
8	519591	523649	4058	0480	Type VI secretion protein
8	519591	523649	4058	0481	Type IV secretion system protein B3, putative
8	519591	523649	4058	0482	Putative secreted protein
8	519591	523649	4058	0483	Type IV secretory pathway, VirB2 component (Pilin)
8	519591	523649	4058	0484	Transposase
8	519591	523649	4058	0485	RND transporter
8	519591	523649	4058	0486	HlyD family secretion protein
9	550787	555370	4583	0514	Uncharacterized protein
9	550787	555370	4583	0515	Putative epoxide hydrolase
9	550787	555370	4583	0516	TonB-dependent receptor
10	564030	568735	4705	0526	PRTRC system protein B
10	564030	568735	4705	0527	PRTRC system protein A
10	564030	568735	4705	0528	Thiamine biosynthesis protein ThiF
10	564030	568735	4705	0529	Uncharacterized protein
10	564030	568735	4705	0530	Uncharacterized protein
10	564030	568735	4705	0531	Uncharacterized protein
10	564030	568735	4705	0532	Uncharacterized protein
11	574301	589855	15554	0534	Uncharacterized protein
11	574301	589855	15554	0535	Uncharacterized protein
11	574301	589855	15554	0536	Type I restriction endonuclease subunit M
11	574301	589855	15554	0537	Uncharacterized protein
11	574301	589855	15554	0538	Uncharacterized protein
11	574301	589855	15554	0539	Uncharacterized protein
11	574301	589855	15554	0540	Uncharacterized protein
11	574301	589855	15554	0541	Uncharacterized protein
11	574301	589855	15554	0542	Methyltransferase type 11
11	574301	589855	15554	0543	Uncharacterized protein
11	574301	589855	15554	0544	Uncharacterized protein
11	574301	589855	15554	0545	Uncharacterized protein
11	574301	589855	15554	0546	Uncharacterized protein
11	574301	589855	15554	0547	Uncharacterized protein
11	574301	589855	15554	0548	Uncharacterized protein
11	574301	589855	15554	0549	Uncharacterized protein
11	574301	589855	15554	0550	Uncharacterized protein
11	574301	589855	15554	0551	Uncharacterized protein
11	574301	589855	15554	0552	Uncharacterized protein
11	574301	589855	15554	0553	Uncharacterized protein
11	574301	589855	15554	0554	Uncharacterized protein
12	593196	599665	6469	0559	Uncharacterized protein
12	593196	599665	6469	0560	Uncharacterized protein
12	593196	599665	6469	0561	Prevent-host-death protein
12	593196	599665	6469	0562	Ribonuclease VapC
12	593196	599665	6469	0563	Uncharacterized protein
12	593196	599665	6469	0564	Uncharacterized protein
12	593196	599665	6469	0565	Uncharacterized protein

## Anexos

Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
12	593196	599665	6469	0566	Uncharacterized protein
12	593196	599665	6469	0567	Uncharacterized protein
12	593196	599665	6469	0568	CopG family transcriptional regulator
13	604988	610389	5401	0571	Restriction endonuclease
13	604988	610389	5401	0572	Lytic transglycosylase catalytic
13	604988	610389	5401	0573	PEP synthase
13	604988	610389	5401	0574	UspA domain-containing protein
14	619606	627568	7962	0577	Aminoglycoside phosphotransferase
14	619606	627568	7962	0578	HspC2 heat shock protein
14	619606	627568	7962	0579	Sensor protein FixL
14	619606	627568	7962	0580	Transcriptional regulatory protein FixJ
14	619606	627568	7962	0581	Putative response regulator
14	619606	627568	7962	0582	Globin-like protein
14	619606	627568	7962	0583	Zinc-binding alcohol dehydrogenase family protein
14	619606	627568	7962	0584	UspA domain-containing protein
15	684134	688268	4134	0639	Cytochrome B561
15	684134	688268	4134	0640	Membrane protein
15	684134	688268	4134	0641	Uncharacterized protein
15	684134	688268	4134	0642	Uncharacterized protein
15	684134	688268	4134	0643	ErfK/YbiS/YcfS/YnhG family protein
16	707273	711582	4309	0659	Putative LuxR family transcriptional regulator
16	707273	711582	4309	0660	Protein of unknown function (DUF2893)
16	707273	711582	4309	0661	Uncharacterized protein (Ync)
16	707273	711582	4309	0662	Uncharacterized protein
16	707273	711582	4309	0663	Uncharacterized protein
16	707273	711582	4309	0664	DNA transfer protein
17	1432460	1447308	14848	1316	Dihydropteroate synthase type-2
17	1432460	1447308	14848	1317	Phosphoglucosamine mutase
17	1432460	1447308	14848	1318	Transposase InsA
17	1432460	1447308	14848	1319	Uncharacterized protein
17	1432460	1447308	14848	1320	Plasmid partitioning protein ParA
17	1432460	1447308	14848	1321	ArsR family transcriptional regulator
17	1432460	1447308	14848	1322	NADPH-dependent FMN reductase Arsh
17	1432460	1447308	14848	1323	Arsenate reductase
17	1432460	1447308	14848	1324	Arsenical-resistance protein
17	1432460	1447308	14848	1325	Putative transcriptional regulator
17	1432460	1447308	14848	1327	Conjugal transfer protein TrbL
17	1432460	1447308	14848	1328	Entry exclusion lipoprotein TrbK
17	1432460	1447308	14848	1329	Conjugal transfer protein TrbJ
17	1432460	1447308	14848	1331	Replication protein C
17	1432460	1447308	14848	1332	RepA replication protein
17	1432460	1447308	14848	1333	Integrase
18	1560087	1565535	5448	1435	Uncharacterized protein
18	1560087	1565535	5448	1436	p pilus assembly protein, porin PapC
18	1560087	1565535	5448	1437	Uncharacterized protein
18	1560087	1565535	5448	1438	Uncharacterized protein
19	1691609	1722457	30848	1557	Integrase
19	1691609	1722457	30848	1558	Uncharacterized protein
19	1691609	1722457	30848	1559	Nuclease
19	1691609	1722457	30848	1560	Lytic transglycosylase
19	1691609	1722457	30848	1561	Uncharacterized protein
19	1691609	1722457	30848	1562	Phage transcriptional regulator, AlpA

Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
19	1691609	1722457	30848	1563	Uncharacterized protein
19	1691609	1722457	30848	1566	Long-chain fatty acid--CoA ligase
19	1691609	1722457	30848	1567	Transposase
19	1691609	1722457	30848	1568	IS511, transposase OrfB
19	1691609	1722457	30848	1569	Thioesterase
19	1691609	1722457	30848	1570	Peroxidase
19	1691609	1722457	30848	1571	Marr family regulator
19	1691609	1722457	30848	1572	Ubiquinone biosynthesis protein
19	1691609	1722457	30848	1573	NUDIX hydrolase
19	1691609	1722457	30848	1574	NADPH:quinone oxidoreductase
19	1691609	1722457	30848	1575	Extradiol ring-cleavage dioxygenase
19	1691609	1722457	30848	1576	Protocatechuate 3,4-dioxygenase
19	1691609	1722457	30848	1577	Ring-hydroxylating dioxygenase large subunit
19	1691609	1722457	30848	1578	3-phenylpropionate dioxygenase subunit beta
19	1691609	1722457	30848	1579	4-hydroxy-2-keto-pentanoic acid aldolase
19	1691609	1722457	30848	1580	Acetaldehyde dehydrogenase 2/3
19	1691609	1722457	30848	1581	2-keto-4-pentenoate hydratase
19	1691609	1722457	30848	1582	Pca regulon regulatory protein
19	1691609	1722457	30848	1583	2,3-dihydroxy-2,3-dihydrophenylpropionate dehydrogenase
19	1691609	1722457	30848	1584	2-hydroxy-6-oxo-2,4-heptadienoate hydrolase
19	1691609	1722457	30848	1585	AMP-dependent synthetase and ligase
19	1691609	1722457	30848	1586	TonB-dependent receptor
19	1691609	1722457	30848	1587	Ferredoxin--NAD(P)(+) reductase fdr
20	1723390	1731543	8153	1589	Transposase
20	1723390	1731543	8153	1590	Mobile element protein
20	1723390	1731543	8153	1591	Integrase catalytic region
20	1723390	1731543	8153	1592	Transposase
20	1723390	1731543	8153	1593	TonB-dependent receptor
21	1734079	1761611	27532	1596	2Fe-2S ferredoxin
21	1734079	1761611	27532	1597	Cytochrome P450
21	1734079	1761611	27532	1598	Regulatory protein GntR, HTH
21	1734079	1761611	27532	1599	Oxidoreductase, 2-nitropropane dioxygenase family protein
21	1734079	1761611	27532	1600	Uncharacterized protein
21	1734079	1761611	27532	1601	Putative nonspecific lipid-transfer protein
21	1734079	1761611	27532	1602	Putative nucleic-acid-binding protein containing a Zn-ribbon
21	1734079	1761611	27532	1603	Putative acyl-CoA dehydrogenase family protein
21	1734079	1761611	27532	1604	Putative acyl-CoA dehydrogenase
21	1734079	1761611	27532	1605	3-oxocholest-4-en-26-oate--CoA ligase
21	1734079	1761611	27532	1606	Cyclohexanone monooxygenase
21	1734079	1761611	27532	1607	Hydrolase, putative
21	1734079	1761611	27532	1608	TonB-dependent receptor
21	1734079	1761611	27532	1609	RND transporter
21	1734079	1761611	27532	1610	DSBA oxidoreductase
21	1734079	1761611	27532	1611	Multidrug ABC transporter permease
21	1734079	1761611	27532	1612	ABC-type transport system periplasmic component
21	1734079	1761611	27532	1613	Uncharacterized protein
21	1734079	1761611	27532	1614	Putative secreted protein
22	1878882	1891309	12427	1715	Integrase
22	1878882	1891309	12427	1716	Lytic transglycosylase
22	1878882	1891309	12427	1717	Uncharacterized protein
22	1878882	1891309	12427	1718	Putative methylase/helicase
22	1878882	1891309	12427	1719	Phage integrase

## Anexos

Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
22	1878882	1891309	12427	1720	Phage integrase
22	1878882	1891309	12427	1721	Integrase
22	1878882	1891309	12427	1722	Putative methylase/helicase
22	1878882	1891309	12427	1723	Uncharacterized protein
22	1878882	1891309	12427	1724	RES domain protein
23	1919110	1924114	5004	1751	Uncharacterized protein
23	1919110	1924114	5004	1752	Thioredoxin-like protein
23	1919110	1924114	5004	1753	Addiction module antitoxin
23	1919110	1924114	5004	1754	Plasmid stabilization protein
23	1919110	1924114	5004	1755	Type IV secretory protease
23	1919110	1924114	5004	1756	hypothetical protein
23	1919110	1924114	5004	1757	Competence protein
23	1919110	1924114	5004	1758	Uncharacterized protein
24	2000239	2009439	9200	1832	Uncharacterized protein
24	2000239	2009439	9200	1833	Uncharacterized protein
24	2000239	2009439	9200	1834	Uncharacterized protein
24	2000239	2009439	9200	1835	Uncharacterized protein
24	2000239	2009439	9200	1836	HNH endonuclease
24	2000239	2009439	9200	1837	Terminase
24	2000239	2009439	9200	1838	Integrase
25	2025209	2032500	7291	1857	Helix-turn-helix domain protein
25	2025209	2032500	7291	1858	LysR family transcriptional regulator
25	2025209	2032500	7291	1859	Epoxide hydrolase domain protein
25	2025209	2032500	7291	1860	LysR family transcriptional regulator
25	2025209	2032500	7291	1861	AraC-type DNA-binding domain-containing protein
25	2025209	2032500	7291	1862	inositol-1-monophosphatase
25	2025209	2032500	7291	1863	Oxidoreductase
25	2025209	2032500	7291	1864	Pentapeptide repeat protein
26	2048540	2055671	7131	1884	Uncharacterized protein
26	2048540	2055671	7131	1885	Filamentation induced by cAMP protein Fic
26	2048540	2055671	7131	1886	Sir2-family regulator
26	2048540	2055671	7131	1887	Nuclease PIN
26	2048540	2055671	7131	1888	Retron-type RNA-directed DNA polymerase
27	2442400	2447644	5244	2245	ATPase
27	2442400	2447644	5244	2246	Uncharacterized protein
27	2442400	2447644	5244	2247	hypothetical protein
28	2569535	2589714	20179	2366	UDP-N-acetylgalactosamine-undecaprenyl-phosphate N-acetylgalactosaminophosphotransferase
28	2569535	2589714	20179	2367	Glycosyl transferase family 1
28	2569535	2589714	20179	2368	Asparagine synthase
28	2569535	2589714	20179	2369	Glycosyl transferase group 1
28	2569535	2589714	20179	2370	Uncharacterized protein
28	2569535	2589714	20179	2371	Glycosyltransferase
28	2569535	2589714	20179	2372	Glycosyl transferase
28	2569535	2589714	20179	2373	FkbM family methyltransferase
28	2569535	2589714	20179	2374	Uncharacterized protein
28	2569535	2589714	20179	2375	Vi polysaccharide biosynthesis protein VipA/TviB
28	2569535	2589714	20179	2376	UDP-glucuronate 5-epimerase
28	2569535	2589714	20179	2377	Uncharacterized protein
28	2569535	2589714	20179	2378	Non-specific protein-tyrosine kinase
28	2569535	2589714	20179	2379	Polysaccharide export protein
28	2569535	2589714	20179	2380	O-antigen polymerase
28	2569535	2589714	20179	2381	Metallophosphoesterase

Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
29	3027481	3038527	11046	2788	TonB-dependent receptor-like protein
29	3027481	3038527	11046	2789	Enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase-like protein
29	3027481	3038527	11046	2790	Pimeloyl-CoA dehydrogenase small subunit-like protein
29	3027481	3038527	11046	2791	Pimeloyl-CoA dehydrogenase large subunit-like protein
29	3027481	3038527	11046	2792	Acetyl-CoA acetyl transferase-like protein
29	3027481	3038527	11046	2793	CoA-transferase-like protein
29	3027481	3038527	11046	2794	Non-acylating NAD-dependent aldehyde dehydrogenase
30	3043041	3060895	17854	2800	Cis-dihydrodiol dehydrogenase
30	3043041	3060895	17854	2801	1,2-dihydroxynaphthalene dioxygenase ThnC
30	3043041	3060895	17854	2802	Rieske-type ferredoxin
30	3043041	3060895	17854	2803	Ferredoxin reductase
30	3043041	3060895	17854	2804	LysR-type activator
30	3043041	3060895	17854	2805	Ferredoxin reductase-like
30	3043041	3060895	17854	2806	Copper resistance protein B
30	3043041	3060895	17854	2807	Copper resistance protein A
30	3043041	3060895	17854	2808	RNA polymerase
30	3043041	3060895	17854	2809	Heavy metal resistance protein
30	3043041	3060895	17854	2810	Uncharacterized protein
30	3043041	3060895	17854	2811	Uncharacterized protein
30	3043041	3060895	17854	2812	Metal-binding protein
30	3043041	3060895	17854	2813	Uncharacterized protein
30	3043041	3060895	17854	2814	Copper-sensing transcriptional repressor CsoR family protein
30	3043041	3060895	17854	2815	Uncharacterized protein
30	3043041	3060895	17854	2816	mRNA 3'-end processing factor
30	3043041	3060895	17854	2817	Putative thymidine phosphorylase
30	3043041	3060895	17854	2818	Ribose-phosphate pyrophosphokinase
30	3043041	3060895	17854	2819	Cu(+)-exporting ATPase
31	3063300	3079834	16534	2820	Membrane protein
31	3063300	3079834	16534	2821	Isoprenylcysteine carboxyl methyltransferase
31	3063300	3079834	16534	2822	Amino acid permease
31	3063300	3079834	16534	2823	Membrane protein
31	3063300	3079834	16534	2824	Uncharacterized protein
31	3063300	3079834	16534	2825	Uncharacterized protein
31	3063300	3079834	16534	2826	Uncharacterized protein
31	3063300	3079834	16534	2827	Uncharacterized protein
31	3063300	3079834	16534	2828	Uncharacterized protein
31	3063300	3079834	16534	2829	Cation transporter
31	3063300	3079834	16534	2830	Cation efflux system protein CusB
31	3063300	3079834	16534	2831	Heavy metal RND transporter
31	3063300	3079834	16534	2832	Methyltransferase type 12
31	3063300	3079834	16534	2834	Cation transporter
31	3063300	3079834	16534	2835	Putative transcriptional regulator
31	3063300	3079834	16534	2836	Uncharacterized protein
31	3063300	3079834	16534	2837	Membrane protein
31	3063300	3079834	16534	2838	Conjugal transfer pilus assembly protein TraL
31	3063300	3079834	16534	2839	Conjugal transfer pilus assembly protein TraE
31	3063300	3079834	16534	2840	Conjugal transfer protein TraK
31	3063300	3079834	16534	2841	Conjugal transfer protein TraB
32	3093114	3102991	9877	2854	Conjugal transfer protein TraG
32	3093114	3102991	9877	2855	Superfamily I DNA/RNA helicase
32	3093114	3102991	9877	2856	Uncharacterized protein

## Anexos

Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
32	3093114	3102991	9877	2857	Alkaline phosphatase
32	3093114	3102991	9877	2858	Helicase
32	3093114	3102991	9877	2859	Putative transposase
32	3093114	3102991	9877	2860	Isocitrate lyase
33	3107612	3127918	20306	2864	Plasmid pRiA4b ORF-3-like family protein
33	3107612	3127918	20306	2865	Helicase domain protein
33	3107612	3127918	20306	2866	DNA methylase
33	3107612	3127918	20306	2867	Transposase IS3/IS911 family protein
33	3107612	3127918	20306	2868	Trnasposase
33	3107612	3127918	20306	2869	DNA methylase
33	3107612	3127918	20306	2870	ATPase
33	3107612	3127918	20306	2871	tRNA delta(2)-isopentenylpyrophosphate transferase
33	3107612	3127918	20306	2872	Transcriptional regulator
33	3107612	3127918	20306	2873	Putative phage repressor
33	3107612	3127918	20306	2874	Uncharacterized protein
33	3107612	3127918	20306	2875	Uncharacterized protein
33	3107612	3127918	20306	2876	Multifunctional conjugation protein TraI
33	3107612	3127918	20306	2877	Type IV secretory pathway VirD4 components-like protein
33	3107612	3127918	20306	2878	Uncharacterized protein
33	3107612	3127918	20306	2879	Uncharacterized protein
33	3107612	3127918	20306	2880	Uncharacterized protein
33	3107612	3127918	20306	2881	Uncharacterized protein
34	3128726	3138976	10250	2883	Uncharacterized protein
34	3128726	3138976	10250	2884	Uncharacterized protein
34	3128726	3138976	10250	2885	Chromosome segregation ATPase
34	3128726	3138976	10250	2886	Uncharacterized protein
34	3128726	3138976	10250	2887	Antirestriction protein ArdC
34	3128726	3138976	10250	2888	Antirestriction protein
34	3128726	3138976	10250	2889	Phage integrase family protein
34	3128726	3138976	10250	2890	Uncharacterized protein
34	3128726	3138976	10250	2891	Putative phage type integrase/recombinase
34	3128726	3138976	10250	2892	Antirestriction protein
35	3141481	3162313	20832	2894	Uncharacterized protein
35	3141481	3162313	20832	2895	DEAD/DEAH box helicase
35	3141481	3162313	20832	2896	hypothetical protein
35	3141481	3162313	20832	2897	Uncharacterized protein
35	3141481	3162313	20832	2898	RES domain protein
35	3141481	3162313	20832	2899	Uncharacterized protein
35	3141481	3162313	20832	2900	Putative methylase/helicase
35	3141481	3162313	20832	2901	Phage integrase
35	3141481	3162313	20832	2902	Phage integrase
35	3141481	3162313	20832	2903	Phage integrase
35	3141481	3162313	20832	2904	Integrase/recombinase
35	3141481	3162313	20832	2905	Integrase/recombinase
35	3141481	3162313	20832	2906	Putative integrase/recombinase y4gC
35	3141481	3162313	20832	2907	Putative methylase/helicase
35	3141481	3162313	20832	2908	Virulence-associated protein E
35	3141481	3162313	20832	2909	Lytic transglycosylase
35	3141481	3162313	20832	2910	Phage integrase
36	3414906	3423254	8348	3113	CRISPR-associated endonuclease Cas1
36	3414906	3423254	8348	3114	Uncharacterized protein
36	3414906	3423254	8348	3115	Uncharacterized protein

Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
36	3414906	3423254	8348	3116	D12 class N6 adenine-specific DNA methyltransferase
36	3414906	3423254	8348	3117	Uncharacterized protein
37	3502309	3506773	4464	3197	Uncharacterized protein
37	3502309	3506773	4464	3198	Uncharacterized protein
38	3516294	3525246	8952	3207	Uncharacterized protein
38	3516294	3525246	8952	3208	Uncharacterized protein
38	3516294	3525246	8952	3209	Putative nucleotidyltransferase
38	3516294	3525246	8952	3210	Transposase
38	3516294	3525246	8952	3211	Transposase
38	3516294	3525246	8952	3212	Integrase
38	3516294	3525246	8952	3213	Nucleotidyltransferase
38	3516294	3525246	8952	3214	Uncharacterized protein
38	3516294	3525246	8952	3215	Uncharacterized protein
38	3516294	3525246	8952	3216	Chemotaxis protein
38	3516294	3525246	8952	3217	Putative transposase
38	3516294	3525246	8952	3218	Isocitrate lyase
39	3540618	3545271	4653	3231	LuxR family transcriptional regulator
39	3540618	3545271	4653	3232	Uncharacterized protein
39	3540618	3545271	4653	3233	Uncharacterized protein
40	3765327	3770696	5369	3447	YD repeat-containing protein
40	3765327	3770696	5369	3448	RHS Repeat protein
40	3765327	3770696	5369	3449	Uncharacterized protein
40	3765327	3770696	5369	3450	Uncharacterized protein
41	4378659	4382908	4249	3981	Phage integrase
41	4378659	4382908	4249	3982	RNA-directed DNA polymerase
41	4378659	4382908	4249	3983	Resolvase domain protein
42	4580571	4584860	4289	4167	PepSY-associated TM helix
42	4580571	4584860	4289	4168	Copper resistance protein B
42	4580571	4584860	4289	4169	Copper resistance protein A
43	4588116	4600864	12748	4170	Cu(+)-exporting ATPase
43	4588116	4600864	12748	4171	HTH-type transcriptional regulator HmrR
43	4588116	4600864	12748	4172	Copper resistance protein D
43	4588116	4600864	12748	4173	Transposase
43	4588116	4600864	12748	4174	Transposase
43	4588116	4600864	12748	4175	Copper resistance protein CopC
43	4588116	4600864	12748	4176	Copper resistance protein A
43	4588116	4600864	12748	4177	Copper resistance protein B
43	4588116	4600864	12748	4178	Cytochrome c family protein
43	4588116	4600864	12748	4179	Uncharacterized protein
43	4588116	4600864	12748	4180	Heavy metal RND efflux outer membrane protein, CzcC family
43	4588116	4600864	12748	4181	Cation efflux system protein CusBr
43	4588116	4600864	12748	4182	Cation transporter
44	4604001	4614202	10201	4183	Copper-binding protein CusF
44	4604001	4614202	10201	4184	DMT(Drug/metabolite transporter) superfamily permease
44	4604001	4614202	10201	4185	Heavy metal translocating P-type ATPase
44	4604001	4614202	10201	4186	MerR family transcriptional regulator
44	4604001	4614202	10201	4187	Iron permease
44	4604001	4614202	10201	4188	Cobalt transporter, cation efflux protein
44	4604001	4614202	10201	4189	Outer membrane efflux protein
44	4604001	4614202	10201	4190	Metal transporter
45	4617463	4638344	20881	4192	Uncharacterized protein
45	4617463	4638344	20881	4193	Heavy metal translocating P-type ATPase

## Anexos

Nº	Inicio	Fin	Tamaño	Locus SGRAN	Producto
45	4617463	4638344	20881	4194	Type I secretion outer membrane protein, TolC
45	4617463	4638344	20881	4195	Uncharacterized protein
45	4617463	4638344	20881	4196	Cadmium, cobalt and zinc/H(+)-K(+) antiporter
45	4617463	4638344	20881	4197	Uncharacterized protein
45	4617463	4638344	20881	4198	Uncharacterized protein
45	4617463	4638344	20881	4199	Uncharacterized protein
45	4617463	4638344	20881	4200	Mercury transporter MerT
45	4617463	4638344	20881	4201	MerP protein
45	4617463	4638344	20881	4202	Hg(II) reductase
45	4617463	4638344	20881	4203	Cation transporter
45	4617463	4638344	20881	4204	Secretion protein HlyD
45	4617463	4638344	20881	4205	Outer membrane efflux protein
45	4617463	4638344	20881	4206	NADPH-dependent FMN reductase ArsH
45	4617463	4638344	20881	4207	Arsenical-resistance protein
45	4617463	4638344	20881	4208	Arsenate reductase
45	4617463	4638344	20881	4209	ArsR family transcriptional regulator
46	4652653	4656862	4209	4219	Uncharacterized protein
46	4652653	4656862	4209	4220	Uncharacterized protein
46	4652653	4656862	4209	4221	Uncharacterized protein
46	4652653	4656862	4209	4222	Uncharacterized protein
46	4652653	4656862	4209	4223	Transcriptional regulator
46	4652653	4656862	4209	4224	RepA replication protein
46	4652653	4656862	4209	4225	Plasmid partitioning protein ParA
46	4652653	4656862	4209	4226	Uncharacterized protein
46	4652653	4656862	4209	4227	Glycosidase
47	4674319	4679853	5534	4244	Conjugal transfer protein Tral
47	4674319	4679853	5534	4245	Uncharacterized protein
47	4674319	4679853	5534	4246	Uncharacterized protein
47	4674319	4679853	5534	4247	Transcriptional regulator, LysR family
47	4674319	4679853	5534	4248	Uncharacterized protein
47	4674319	4679853	5534	4249	Abortive phage resistance protein
47	4674319	4679853	5534	4250	Uncharacterized protein



**Anexo V. Genes específicos de TFA dentro del género *Sphingopyxis***

Locus SGRAN	Inicio	Fin	Cadena	Gen	Descripción
0032	40390	39014	-		UBA/THIF-type NAD/FAD binding protein
0034	41207	40794	-		Uncharacterized protein
0035	41360	41722	+		Uncharacterized protein
0036	41722	42645	+		Uncharacterized protein
0037	42649	42999	+		Uncharacterized protein
0038	43478	44722	+		Uncharacterized protein
0039	44715	45341	+		dCTP deaminase
0043	48748	48497	-		Uncharacterized protein
0045	50766	49492	-		Phosphatidylinositol kinase
0046	51106	50756	-		XRE family transcriptional regulator
0136	145624	145779	+		Uncharacterized protein
0149	162604	162074	-		Uncharacterized protein
0150	163750	162683	-		Putative uncharacterized protein
0151	166371	163747	-		Uncharacterized protein
0152	167561	166437	-		Uncharacterized protein
0154	168110	168868	+		ETC complex I subunit region
0155	169553	169825	+		Uncharacterized protein
0156	170083	169844	-		Hydratase/decarboxylase family protein
0157	170940	172085	+		Carbamoyl-phosphate synthase, large subunit
0158	172150	172461	+		Uncharacterized protein
0160	173796	173990	+		Uncharacterized protein
0163	176608	176784	+		Uncharacterized protein
0198	207662	207979	+		Uncharacterized protein
0203	216097	215681	-		Uncharacterized protein
0204	216718	216161	-		Uncharacterized protein
0205	217048	216827	-		Uncharacterized protein
0214	227622	228218	+		Uncharacterized protein
0216	229244	230026	+		Uncharacterized protein
0218	233704	234558	+		Uncharacterized protein
0219	234555	235133	+		Uncharacterized protein
0221	238618	239364	+		Putative uncharacterized protein
0222	239339	240511	+		ThiF family protein
0233	254905	253928	-		Putative MarR family transcriptional regulator
0471	505411	501128	-		YD repeat protein
0472	506127	505408	-		Hemagglutinin-related protein
0473	506368	506964	+		Uncharacterized protein
0474	507835	512004	+		ParB-like nuclease domain
0493	530649	529684	-		SnoAL-like domain protein
0495	532806	532315	-		Acyl-CoA carboxylase biotin/lipoyl carrier
0506	544067	543453	-		MarR family regulator
0508	546153	544939	-		Acyl-CoA dehydrogenase protein
0514	550787	551590	+		Uncharacterized protein
0519	558381	557998	-		Uncharacterized protein
0520	558920	559267	+		Regulatory protein MarR

## Anexos

Locus SGRAN	Inicio	Fin	Cadena	Gen	Descripción
0522	560468	562117	+		Chromosome partitioning protein ParB
0523	562155	562592	+		PRTRC system protein E
0524	562642	562848	+		Uncharacterized protein
0525	563236	564018	+		Uncharacterized protein
0526	564030	564860	+		PRTRC system protein B
0527	564857	565516	+		PRTRC system protein A
0528	565506	566264	+		Thiamine biosynthesis protein ThiF
0529	566379	566654	+		Uncharacterized protein
0530	566651	566935	+		Uncharacterized protein
0532	567497	568735	+		Uncharacterized protein
0533	568856	574000	+		Lactate dehydrogenase
0534	574301	574777	+		Uncharacterized protein
0535	575034	576281	+		Uncharacterized protein
0536	576449	577258	+		Type I restriction endonuclease subunit M
0540	579330	580085	+		Uncharacterized protein
0542	581096	582577	+		Methyltransferase type 11
0543	582574	583092	+		Uncharacterized protein
0546	585460	586050	+		Uncharacterized protein
0548	586630	586995	+		Uncharacterized protein
0550	587498	587713	+		Uncharacterized protein
0552	587998	588192	+		Uncharacterized protein
0553	588189	588737	+		Uncharacterized protein
0554	589454	589855	+		Uncharacterized protein
0556	590566	591651	+		Uncharacterized protein
0557	591788	591988	+		Uncharacterized protein
0558	592080	593084	+		Uncharacterized protein
0559	593372	593196	-		Uncharacterized protein
0560	593607	594209	+		Uncharacterized protein
0571	604988	605578	+		Restriction endonuclease
0573	606698	609079	+	ppsA	PEP synthase
0574	609376	610389	+		UspA domain-containing protein
0575	610393	618804	+		Carbohydrate binding protein
0576	618916	619152	+		Uncharacterized protein
0577	621066	619606	-		Aminoglycoside phosphotransferase
0582	625170	625580	+		Globin-like protein
0589	632117	634822	+	ctpF	Carbonate dehydratase
0592	636945	637874	+	pfkB	Phosphofructokinase
0601	646199	646825	+		Glutamate synthase domain-containing 2
0611	656238	655822	-		Uncharacterized protein
0622	667511	667738	+		TraR/DksA family transcriptional regulator
0630	673816	673100	-		TonB family protein
0631	676057	673823	-		Putative TonB-dependent receptor
0634	678121	677822	-		Uncharacterized protein
0639	684823	684134	-		Cytochrome B561
0641	686042	685791	-		Uncharacterized protein

Locus SGRAN	Inicio	Fin	Cadena	Gen	Descripción
0642	687187	686537	-		Uncharacterized protein
0645	689433	690935	+	cls	Cardiolipin synthase
0646	691604	691020	-		Uncharacterized protein
0648	692719	692147	-		Putative membrane protein
0649	694083	692860	-		Putative outer membrane phospholipase A
0650	695273	694080	-	ackA	Acetate kinase
0651	696691	695270	-	pta	Phosphotransbutyrylase
0654	700528	699704	-		Putative TetR family transcriptional regulator
0657	703102	705852	+	rbbA	Multidrug ABC transporter ATP-binding protein
0658	705854	706978	+	yhhJ	ABC-2 type transporter
0660	708258	709043	+		Protein of unknown function (DUF2893)
0661	709048	709968	+		Uncharacterized protein (Ync)
0662	710535	710182	-		Uncharacterized protein
0675	725502	724987	-		Uncharacterized protein
0676	725903	725493	-		Uncharacterized protein
0686	734064	733759	-		Uncharacterized protein
0688	734979	735647	+	traA	Conjugal transfer protein TraA
0689	736556	735855	-	pltI	Secretion protein HlyD family protein
0690	736660	737343	+	pltZ	Transcriptional regulator PltZ
0691	737324	738565	+		Phospholipase A
0692	738924	739127	+		Excisionase
0693	739088	740224	+		Putative site-specific recombinase
0860	921774	922610	+		Uncharacterized protein
1037	1113486	1112569	-		Peptidase S15
1076	1159038	1161356	+		Acyl-CoA dehydrogenase
1258	1369955	1369680	-		Putative uncharacterized protein
1259	1370786	1370178	-		Uncharacterized protein
1317	1433362	1433664	+	glmM	Phosphoglucosamine mutase
1318	1435333	1433840	-	insA	Transposase InsA
1319	1435768	1435544	-		Uncharacterized protein
1325	1439428	1439631	+		Putative transcriptional regulator
1327	1441636	1440209	-	trbL2	Conjugal transfer protein TrbL
1328	1441883	1441647	-		Entry exclusion lipoprotein TrbK
1329	1442675	1441896	-	trbJ2	Conjugal transfer protein TrbJ
1331	1444918	1444037	-	repC	Replication protein C
1332	1445732	1444905	-	repA2	RepA replication protein
1430	1553158	1552055	-		TonB-dependent receptor
1433	1556352	1554769	-		TonB-dependent receptor
1437	1563413	1564300	+		Uncharacterized protein
1449	1578061	1575854	-		Uncharacterized protein
1453	1581067	1580033	-		Uncharacterized protein
1498	1631606	1631304	-		Uncharacterized protein
1508	1639998	1640117	+		Transposase for insertion sequence element
1510	1641549	1641019	-		RES domain protein
1511	1641962	1641546	-		XRE family transcriptional regulator

## Anexos

Locus SGRAN	Inicio	Fin	Cadena	Gen	Descripción
1516	1644538	1644227	-	trbF2	Conjugal transfer protein TrbF
1539	1669748	1668741	-		Uncharacterized protein
1559	1693705	1693154	-		Nuclease
1563	1695610	1696170	+		Uncharacterized protein
1566	1698491	1699081	+	alkK3	Long-chain fatty acid--CoA ligase
1575	1707747	1708031	+		Extradiol ring-cleavage dioxygenase
1576	1708089	1708907	+		Protocatechuate 3,4-dioxygenase
1588	1723074	1723172	+		Probable insertion sequence transposase protein
1589	1723390	1724022	+		Transposase
1590	1725967	1725791	-		Mobile element protein
1591	1726820	1727029	+		Integrase catalytic region
1612	1755493	1756443	+		ABC-type transport system periplasmic component
1615	1761858	1764950	+		Uncharacterized protein
1616	1766209	1764860	-		histidine kinase
1657	1823942	1822551	-		Transposase
1705	1868676	1868422	-		Uncharacterized protein
1736	1902675	1902917	+	hlyB	Type I secretion system ATPase
1737	1903029	1903307	+	hlyB2	Hemolysin secretion protein HlyB
1743	1911302	1910838	-	fcB	Thioesterase
1745	1914357	1912186	-		TonB-dependent receptor
1747	1917323	1916691	-	traH2	Conjugal transfer protein TraH
1750	1919056	1918616	-	traH3	Conjugal transfer protein TraH
1754	1921120	1921449	+		Plasmid stabilization protein
1755	1921676	1921512	-		Type IV secretory protease
1756	1921828	1922133	+		Putative uncharacterized protein
1757	1922240	1923286	+		Competence protein
1758	1923374	1924114	+		Uncharacterized protein
1764	1929960	1929640	-	nnrD	ADP-dependent NAD(P)HX dehydratase
1783	1951037	1951312	+	luxR	Transcriptional regulator, LuxR family
1832	2001300	2000239	-		Uncharacterized protein
1833	2001817	2001302	-		Uncharacterized protein
1834	2002422	2001814	-		Uncharacterized protein
1835	2003687	2002422	-		Uncharacterized protein
1836	2004009	2003698	-		HNH endonuclease
1837	2005517	2004006	-		Terminase
1838	2008279	2009439	+		Integrase
1853	2021770	2022186	+		Uncharacterized protein
1864	2032500	2031880	-		Pentapeptide repeat protein
1866	2035017	2034697	-		Uncharacterized protein
1874	2039718	2039515	-		Uncharacterized protein
1875	2040866	2040024	-		Unnamed protein product
1881	2045024	2045473	+		Uncharacterized protein
1884	2048947	2048540	-		Uncharacterized protein
1886	2050474	2052357	+		Sir2-family regulator
1887	2052357	2053805	+		Nuclease PIN

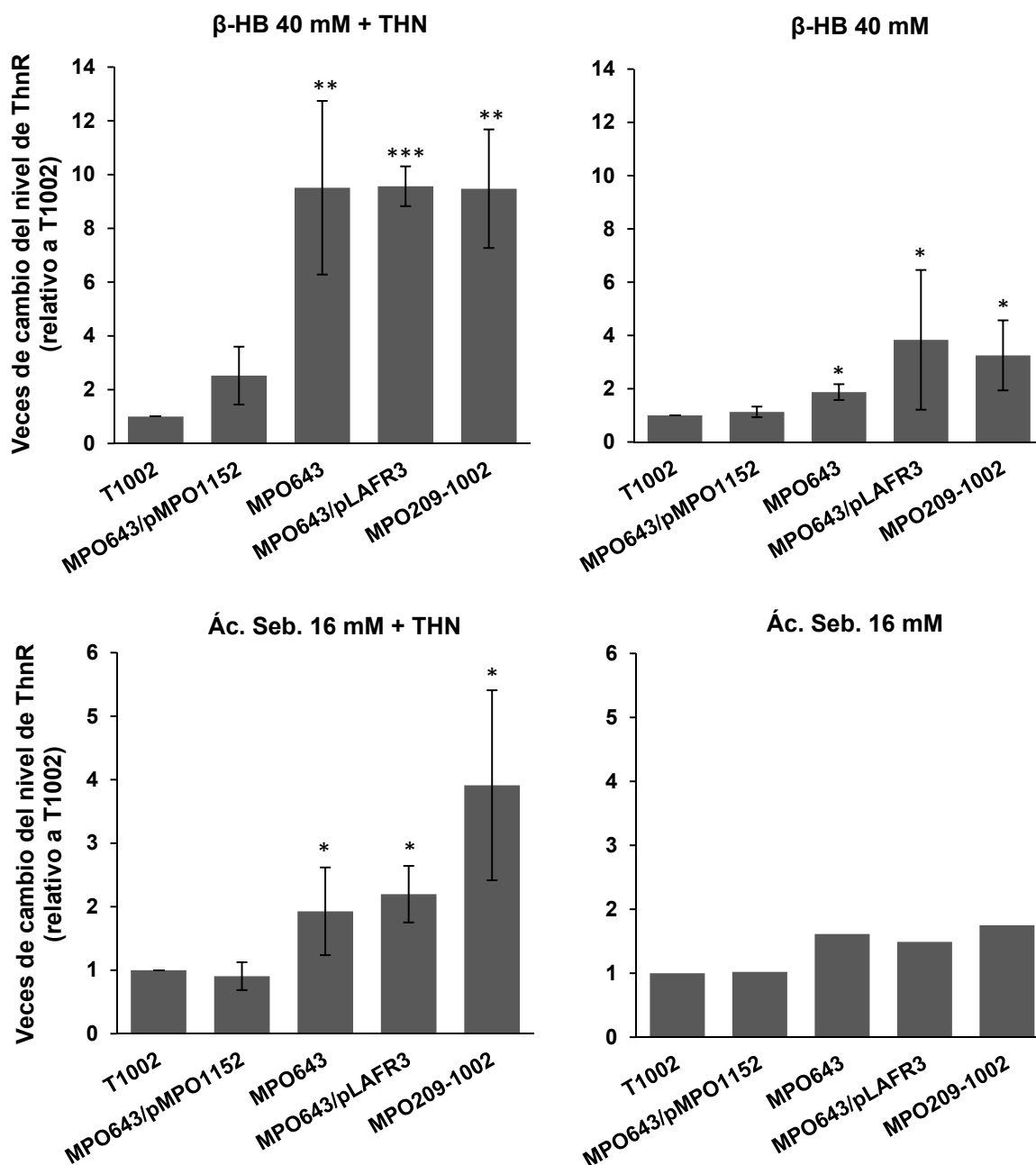
Locus SGRAN	Inicio	Fin	Cadena	Gen	Descripción
1888	2055671	2053929	-		Retron-type RNA-directed DNA polymerase
1978	2153995	2155752	+		Uncharacterized protein
2000	2179480	2179701	+		Pathogenicity locus
2001	2179701	2180234	+		ATPase
2038	2223709	2223882	+		Conserved Putative membrane protein
2196	2382010	2382162	+		Uncharacterized protein
2245	2446809	2442400	-		ATPase
2246	2447644	2447063	-		Uncharacterized protein
2247	2448345	2447641	-		Putative uncharacterized protein
2357	2558283	2557975	-		DNA-binding protein
2358	2558622	2558293	-	relE	RelE cytotoxic translational repressor of toxin-antitoxin stability system
2366	2570878	2569535	-	wecA	UDP-N-acetylgalactosamine-undecaprenyl-phosphate N-acetylgalactosaminephosphotransferase
2370	2576450	2575521	-		Uncharacterized protein
2372	2578529	2577606	-		Glycosyl transferase
2373	2579365	2578532	-		FkbM family methyltransferase
2374	2580639	2579362	-		Uncharacterized protein
2465	2675132	2674149	-		Protein of unassigned function
2500	2708796	2709563	+		Acyl-CoA dehydrogenase
2503	2710859	2711269	+		Acyl-CoA dehydrogenase
2504	2712145	2711642	-	traG4	Conjugal transfer protein TraG
2509	2715611	2715378	-		Uncharacterized protein
2513	2717740	2717522	-		2,5-diketo-D-gluconic acid reductase
2518	2723801	2723649	-		Uncharacterized protein
2536	2742824	2743114	+	parD1	Addiction module antidote protein
2537	2743114	2743491	+		Plasmid stabilization system protein
2538	2744656	2743778	-		Uncharacterized protein
2548	2756357	2754501	-	ltrA1	RNA-directed DNA polymerase
2549	2757733	2756357	-	ardC4	Antirepressor
2550	2758255	2757803	-		Uncharacterized protein
2551	2759584	2759234	-		Uncharacterized protein
2570	2773720	2774874	+		Uncharacterized protein
2607	2814481	2814374	-		Putative bacteriochlorophyll 4-vinyl reductase
2610	2819168	2818101	-		TonB-dependent receptor
2672	2884062	2884925	+		Peptidase S15
2673	2884943	2886214	+	fadB	3-hydroxyacyl-CoA dehydrogenase
2675	2890091	2887617	-		TonB-dependent receptor
2681	2897399	2895414	-	pvdh	PVA dehydrogenase
2682	2898493	2897399	-	oph	Oxidized polyvinyl alcohol hydrolase
2686	2906063	2906197	+		Magnesium transporter
2693	2917024	2915834	-		Major facilitator transporter
2695	2918193	2917507	-		Putative ankyrin-containing lipoprotein
2696	2919395	2918181	-	atzC	Cytosine deaminase
2784	3024262	3020801	-	pyc	Pyruvate carboxylase

## Anexos

Locus SGRAN	Inicio	Fin	Cadena	Gen	Descripción
2788	3029652	3027481	-	thnM	TonB-dependent receptor-like protein
2803	3045585	3046604	+	thnA4	Ferredoxin reductase
2805	3047892	3048683	+	thnY	Ferredoxin reductase-like
2813	3055924	3056307	+		Uncharacterized protein
2820	3063300	3063539	+		Membrane protein
2821	3063536	3064180	+		Isoprenylcysteine carboxyl methyltransferase
2827	3067401	3067787	+		Uncharacterized protein
2832	3075314	3074694	-		Methyltransferase type 12
2845	3085376	3085771	+		Conjugative transfer protein
2855	3097906	3096110	-		Superfamily I DNA/RNA helicase
2856	3098637	3097918	-		Uncharacterized protein
2857	3100627	3098624	-		Alkaline phosphatase
2858	3102345	3100624	-		Helicase
2865	3109158	3108178	-		Helicase domain protein
2866	3110330	3109155	-		DNA methylase
2869	3113247	3112297	-		DNA methylase
2870	3117003	3113278	-		ATPase
2871	3117489	3117031	-		tRNA delta(2)-isopentenylpyrophosphate transferase
2874	3119533	3120168	+		Uncharacterized protein
2878	3126391	3126059	-		Uncharacterized protein
2879	3126831	3127025	+		Uncharacterized protein
2881	3127688	3127918	+		Uncharacterized protein
2894	3141777	3141481	-		Uncharacterized protein
2895	3144400	3141791	-		DEAD/DEAH box helicase
2896	3145359	3144397	-		Putative uncharacterized protein
3113	3415973	3414906	-	cas1	CRISPR-associated endonuclease Cas1
3114	3417028	3417597	+		Uncharacterized protein
3115	3417972	3420977	+		Uncharacterized protein
3116	3421942	3421055	-	dam	Adenine-specific DNA methyltransferase
3117	3423254	3421977	-		Uncharacterized protein
3185	3495150	3496124	+		Uncharacterized protein
3191	3499233	3499024	-		Uncharacterized protein
3195	3501061	3501675	+		Uncharacterized protein
3197	3504747	3502309	-		Uncharacterized protein
3198	3506000	3506773	+		Uncharacterized protein
3207	3516294	3517106	+		Uncharacterized protein
3208	3517484	3518047	+		Uncharacterized protein
3211	3519376	3520041	+		Transposase
3212	3520002	3520244	+		Integrase
3213	3520496	3520263	-		Nucleotidyltransferase
3215	3521254	3522939	+		Uncharacterized protein
3216	3522936	3524741	+		Chemotaxis protein
3221	3528791	3528075	-		Uncharacterized protein
3232	3542157	3544619	+		Uncharacterized protein
3233	3545271	3544627	-		Uncharacterized protein

Locus SGRAN	Inicio	Fin	Cadena	Gen	Descripción
3241	3554992	3553871	-		3,6-diketocamphane 1,6 monooxygenase
3242	3555133	3555786	+		Regulatory protein TetR
3248	3559907	3559446	-		Uncharacterized protein
3250	3562966	3562454	-		Uncharacterized protein
3251	3563069	3563716	+		Uncharacterized protein
3255	3568031	3566931	-	camP	Baeyer-Villiger monooxygenase
3256	3569226	3568072	-		Luciferase
3260	3573473	3573162	-		Amidohydrolase 2
3261	3574531	3573470	-		Amidohydrolase 2
3277	3592063	3593142	+		Butanediol dehydrogenase
3279	3594028	3595185	+		Rieske (2Fe-2S) iron-sulfur domain protein
3280	3595208	3596272	+		Oxidoreductase
3282	3596977	3598056	+	camP2	Baeyer-Villiger monooxygenase
3286	3601015	3601509			Glyoxalase/bleomycin resistance protein
3447	3769652	3765327			YD repeat-containing protein
3718	4066621	4067397	+	tnpC6	Transposase
3765	4119907	4119722	-		Uncharacterized protein
3837	4201083	4200928	-		Multidrug transporter AcrB
3919	4292395	4290713	-		ATPase domain protein
3938	4321880	4321638	-		Uncharacterized protein
3947	4332961	4333227	+		Uncharacterized protein
3981	4379792	4378659	-		Phage integrase
3982	4381168	4382250	+		RNA-directed DNA polymerase
3984	4384805	4384963	+		tetracycline resistance MFS efflux pump
4166	4578539	4580530	+	aleB	TonB-dependent siderophore receptor
4175	4591364	4591056	-	copC	Copper resistance protein CopC
4184	4604421	4605497	+	cnrT	DMT(Drug/metabolite transporter) permease
4188	4610621	4610995	+		Cobalt transporter, cation efflux protein
4195	4622607	4623260	+		Uncharacterized protein
4213	4642532	4643167	+		Uncharacterized protein
4221	4653668	4653865	+		Uncharacterized protein
4232	4661023	4661952	+		Uncharacterized protein
4240	4669545	4669937	+		Putative secreted protein
4247	4675205	4674891	-		Transcriptional regulator, LysR family
4248	4677420	4676434	-		Uncharacterized protein
4250	4679853	4679191	-		Uncharacterized protein

**Anexo VI. Cuantificación normalizada de la señal de los *Western Blot* para detectar los niveles de ThnR**



p-value < 0,5 \*\*p-value < 0,05 \*\*\*p-value < 0,005 (T-student entre los mutantes y T1002).

En la condición de MM con ácido sebáico 16 mM solo se realizó una repetición, en MM con ácido sebáico 16 mM + THN y en MM con β-HB 40 mM dos réplicas y en MM con β-HB 40 mM + THN tres réplicas.



## **BIBLIOGRAFÍA**

---



- Alcaraz, L.D., Moreno-Hagelsieb, G., Eguiarte, L.E., Souza, V., Herrera-Estrella, L. y Olmedo, G. (2010). Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC genomics* 11:332.
- Ali, S., Duan, J., Charles, T.C. y Glick, B.R. (2014). A bioinformatics approach to the determination of genes involved in endophytic behavior in *Burkholderia* spp. *Journal of Theoretical Biology* 343:193-198.
- Alikhan, N., Petty, N.K., Ben Zakour, N.L. y Beatson, S.A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics* 12:402.
- Allocati, N., Federici, L., Masulli, M. and Di Ilio, C. (2009). Glutathione transferases in bacteria. *The FEBS journal* 276:58-75.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. y Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215:403-410.
- Andersen, J., Forst, S.A., Zhao, K., Inouye, M. y Delihias, N. (1989). The function of *micF* RNA. *micF* RNA is a major factor in the thermal regulation of OmpF protein in *Escherichia coli*. *The Journal of Biological Chemistry* 264:17961-17970.
- Andújar, E., Hernández, M.J., Kaschabek, S.R., Reineke, W. y Santero, E. (2000). Identification of an extradiol dioxygenase involved in tetralin biodegradation: gene sequence analysis and purification and characterization of the gene product. *Journal of Bacteriology* 182:789-795.
- Andújar, E. y Santero, E. (2003). Site-directed mutagenesis of an extradiol dioxygenase involved in tetralin biodegradation identifies residues important for activity or substrate specificity. *Microbiology* 149:1559-1567.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M. y col. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* 25:25-29.
- Aylward, F.O., McDonald, B.R., Adams, S.M., Valenzuela, A., Schmidt, R.A., Goodwin, L.A. y col. (2013). Comparison of 26 sphingomonad genomes reveals diverse environmental adaptations and biodegradative capabilities. *Applied and Environmental Microbiology* 79:3724-3733.
- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A. y col. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC genomics* 9:75.
- Babski, J., Haas, K.A., Näther-Schindler, D., Pfeiffer, F., Förstner, K.U., Hammelmann, M. y col. (2016). Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC genomics* 17:629.
- Bak, G., Han, K., Kim, K. y Lee, Y. (2015). Electrophoretic mobility shift assay of RNA-RNA complexes. *Methods in Molecular Biology* 1240:153-163.
- Balkwill, D.L., Fredrickson, J.K. y Romine, M.F. (2006). *Sphingomonas* and Related Genera. En: Dr, M.D.P., Falkow, S., Rosenberg, E., Schleifer, K. y Stackebrandt, E. (Eds.), *The Prokaryotes*, Springer New York. 605-629.
- Basta, T., Buerger, S. y Stolz, A. (2005). Structural and replicative diversity of large plasmids from sphingomonads that degrade polycyclic aromatic compounds and xenobiotics. *Microbiology* 151:2025-2037.

## Bibliografía

- Basta, T., Keck, A., Klein, J. y Stolz, A. (2004). Detection and characterization of conjugative degradative plasmids in xenobiotic-degrading *Sphingomonas* strains. *Journal of Bacteriology* 186:3862-3872.
- Battu, L., Reddy, M.M., Goud, B.S., Ulaganathan, K. y Kandasamy, U. (2017a). Genome inside genome: NGS based identification and assembly of endophytic *Sphingopyxis granuli* and *Pseudomonas aeruginosa* genomes from rice genomic reads. *Genomics* 109:141-146.
- Battu, L., Reddy, M.M., Goud, B.S., Ulaganathan, K. y Ulaganathan, K. (2017b). Assembly of genomic reads of elite indica rice cultivar onto 2101 reference bacterial genomes for identification of co-sequenced endophytic bacteria. *Data in Brief* 12:305-312.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Brice, M.D., Rodgers, J.R. y col. (1977). The Protein Data Bank. A computer-based archival file for macromolecular structures. *European Journal of Biochemistry* 80:319-324.
- Besemer, J., Lomsadze, A. y Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research* 29:2607-2618.
- Binns, N. y Masters, M. (2002). Expression of the *Escherichia coli* *pcnB* gene is translationally limited using an inefficient start codon: a second chromosomal example of translation initiated at AUU. *Molecular Microbiology* 44:1287-1298.
- Bischler, T., Tan, H.S., Nieselt, K. y Sharma, C.M. (2015). Differential RNA-seq (dRNA-seq) for annotation of transcriptional start sites and small RNAs in *Helicobacter pylori*. *Methods* 86:89-101.
- Blencke, H., Homuth, G., Ludwig, H., Mäder, U., Hecker, M. y Stülke, J. (2003). Transcriptional profiling of gene expression in response to glucose in *Bacillus subtilis*: regulation of the central metabolic pathways. *Metabolic Engineering* 5:133-149.
- Boyd, E.F., Almagro-Moreno, S. y Parent, M.A. (2009). Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends in Microbiology* 17:47-53.
- Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. *Current Opinion in Microbiology* 10:102-109.
- Breidenbach, B., Pump, J. y Dumont, M.G. (2015). Microbial Community Structure in the Rhizosphere of Rice Plants. *Frontiers in Microbiology* 6:1537.
- Burse, A., Weingart, H. y Ullrich, M.S. (2004). The phytoalexin-inducible multidrug efflux pump AcrAB contributes to virulence in the fire blight pathogen, *Erwinia amylovora*. *Molecular plant-microbe interactions* 17:43-54.
- Busch, A., Richter, A.S. y Backofen, R. (2008). IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* 24(24):2849-2856.
- Carver, T., Thomson, N., Bleasby, A., Berriman, M. y Parkhill, J. (2009). DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25:119-120.
- Casimiro-Soriguer, C.S., Muñoz-Mérida, A. y Pérez-Pulido, A.J. (2017). Sma3s: A universal tool for easy functional annotation of proteomes and transcriptomes. *Proteomics* 17.

- Ceizel Borella, G., Lagares, A. y Valverde, C. (2016). Expression of the *Sinorhizobium meliloti* small RNA gene *mmgR* is controlled by the nitrogen source. *FEMS microbiology letters* 363.
- Chan, P.P. y Lowe, T.M. (2009). GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Research* 37:93.
- Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C.M. y Vogel, J. (2012). An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *The EMBO journal* 31:4005-4019.
- Chen, B., Zhang, Y., Rafiq, M.T., Khan, K.Y., Pan, F., Yang, X. y col. (2014). Improvement of cadmium uptake and accumulation in *Sedum alfredii* by endophytic bacteria *Sphingomonas* SaMR12: effects on plant growth and root exudates. *Chemosphere* 117:367-373.
- Chen, S., Zhang, A., Blyn, L.B. y Storz, G. (2004). MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*. *Journal of Bacteriology* 186:6689-6697.
- Chiellini, E., Corti, A., D'Antone, S. y Solaro, R. (2003). Biodegradation of poly (vinyl alcohol) based materials. *Progress in Polymer Science* 28:963-1014.
- Choi, D., Lee, S.B., Kim, S., Min, B., Choi, I. y Chang, I.S. (2014). Metabolically engineered glucose-utilizing *Shewanella* strains under anaerobic conditions. *Bioresource Technology* 154:59-66.
- Chung, C.T., Niemela, S.L. y Miller, R.H. (1989). One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences of the United States of America* 86:2172-2175.
- Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P. y Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS microbiology reviews* 29:147-167.
- Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y. y col. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research* 42:633.
- Copley, S.D., Rokicki, J., Turner, P., Daligault, H., Nolan, M. y Land, M. (2012). The whole genome sequence of *Sphingobium chlorophenolicum* L-1: insights into the evolution of the pentachlorophenol degradation pathway. *Genome biology and evolution* 4:184-198.
- Corbino, K.A., Barrick, J.E., Lim, J., Welz, R., Tucker, B.J., Puskarz, I. y col. (2005). Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biology* 6:R70.
- Corcoran, C.P., Podkaminski, D., Papenfort, K., Urban, J.H., Hinton, J.C.D. y Vogel, J. (2012). Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Molecular Microbiology* 84:428-445.
- Coulondre, C., Miller, J.H., Farabaugh, P.J. y Gilbert, W. (1978). Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* 274:775-780.
- Darty, K., Denise, A. y Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25:1974-1975.

## Bibliografía

- D'Auria, G., Jiménez-Hernández, N., Peris-Bondia, F., Moya, A. y Latorre, A. (2010). *Legionella pneumophila* pangenome reveals strain-specific virulence factors. *BMC genomics* 11:181.
- Delcher, A.L., Bratke, K.A., Powers, E.C. y Salzberg, S.L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673-679.
- Dhillon, B.K., Laird, M.R., Shay, J.A., Winsor, G.L., Lo, R., Nizam, F. y col. (2015). IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic acids research* 43:W104-W108.
- Díaz, E., Ferrández, A., Prieto, M.A. y García, J.L. (2001). Biodegradation of Aromatic Compounds by *Escherichia coli*. *Microbiology and Molecular Biology Reviews* 65:523-569.
- Díaz, E., Jiménez, J.I. y Nogales, J. (2013). Aerobic degradation of aromatic compounds. *Current opinion in biotechnology* 24:431-442.
- Dinamarca, M.A., Ruiz-Manzano, A. y Rojo, F. (2002). Inactivation of cytochrome o ubiquinol oxidase relieves catabolic repression of the *Pseudomonas putida* GPo1 alkane degradation pathway. *Journal of Bacteriology* 184:3785-3793.
- Dorn, E., Hellwig, M., Reineke, W. y Knackmuss, H.J. (1974). Isolation and characterization of a 3-chlorobenzoate degrading pseudomonad. *Archives of Microbiology* 99:61-70.
- Dua, M., Singh, A., Sethunathan, N. y Johri, A.K. (2002). Biotechnology and bioremediation: successes and limitations. *Applied Microbiology and Biotechnology* 59:143-152.
- Dugar, G., Herbig, A., Förstner, K.U., Heidrich, N., Reinhardt, R., Nieselt, K. y col. (2013). High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. *PLoS genetics* 9:e1003495.
- Dürring, U., Axmann, I.M., Hess, W.R. y Wilde, A. (2006). An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. *Proceedings of the National Academy of Sciences of the United States of America* 103:7054-7058.
- Dunwell, J.M., Purvis, A. y Khuri, S. (2004). Cupins: the most functionally diverse protein superfamily? *Phytochemistry* 65:7-17.
- Eppinger, M., Baar, C., Raddatz, G., Huson, D.H. y Schuster, S.C. (2004). Comparative analysis of four Campylobacterales. *Nature Reviews Microbiology* 2:872-885.
- Fallet, C., Rohe, P. y Franco-Lara, E. (2010). Process optimization of the integrated synthesis and secretion of ectoine and hydroxyectoine under hyper/hypo-osmotic stress. *Biotechnology and Bioengineering* 107:124-133.
- Fan, B., Li, L., Chao, Y., Förstner, K., Vogel, J., Borriss, R. y col. (2015). dRNA-Seq Reveals Genomewide TSSs and Noncoding RNAs of Plant Beneficial Rhizobacterium *Bacillus amyloliquefaciens* FZB42. *PloS One* 10:e0142002.
- Faucher, S.P. y Shuman, H.A. (2011). Small Regulatory RNA and *Legionella pneumophila*. *Frontiers in Microbiology* 2:98.
- Fegatella, F., Lim, J., Kjelleberg, S. y Cavicchioli, R. (1998). Implications of rRNA operon copy number and ribosome content in the marine oligotrophic ultramicrobacterium

- Sphingomonas* sp. strain RB2256. *Applied and Environmental Microbiology* 64:4433-4438.
- Feist, A.M., Herrgård, M.J., Thiele, I., Reed, J.L. y Palsson, B.Ø. (2009). Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology* 7:129-143.
- Fernández, H., Prandoni, N., Fernández-Pascual, M., Fajardo, S., Morcillo, C., Díaz, E. y col. (2014). *Azoarcus* sp. CIB, an anaerobic biodegrader of aromatic compounds shows an endophytic lifestyle. *PloS One* 9:e110771.
- Ferrante, A.A., Augliera, J., Lewis, K. y Klibanov, A.M. (1995). Cloning of an organic solvent-resistance gene in *Escherichia coli*: the unexpected role of alkylhydroperoxide reductase. *Proceedings of the National Academy of Sciences of the United States of America* 92:7617-7621.
- Figurski, D.H. y Helinski, D.R. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proceedings of the National Academy of Sciences of the United States of America* 76:1648-1652.
- Findeiss, S., Schmidtke, C., Stadler, P.F. y Bonas, U. (2010). A novel family of plasmid-transferred anti-sense ncRNAs. *RNA biology* 7:120-124.
- Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E. y col. (2010). The Pfam protein families database. *Nucleic Acids Research* 38:211.
- Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R. y col. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496-512.
- Fouts, D.E., Tyker, H.L., DeBoy, R.T., Dauguerty, S., Ren, Q., Badger J.H. y col. (2008). Complete Genome Sequence of the N<sub>2</sub>-Fixing Broad Host Range Endophyte *Klebsiella pneumoniae* 342 and Virulence Predictions Verified in Mice. *Plos Genetics* 4:e1000141.
- Fozo, E.M., Hemm, M.R. y Storz, G. (2008). Small toxic proteins and the antisense RNAs that repress them. *Microbiology and molecular biology reviews* 72:589.
- Frank, A.C. y Lobry, J.R. (1999). Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene* 238:65-77.
- Fraser, C.M., Eisen, J., Fleischmann, R.D., Ketchum, K.A. y Peterson, S. (2000). Comparative genomics and understanding of microbial biology. *Emerging Infectious Diseases* 6:505-512.
- Freese, N.H., Norris, D.C. y Loraine, A.E. (2016). Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* 32:2089-2095.
- Fuchs, G., Boll, M. y Heider, J. (2011). Microbial degradation of aromatic compounds - from one strategy to four. *Nature Reviews Microbiology* 9:803-816.
- Gao, F. (2014). Recent Advances in the Identification of Replication Origins Based on the Z-curve Method. *Current Genomics* 15:104-112.
- Gao, F., Luo, H. y Zhang, C. (2013). DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. *Nucleic Acids Research* 41:90.
- Gao, F. y Zhang, C. (2008). Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. *BMC bioinformatics* 9:79.

## Bibliografía

- García, L.L., Rivas-Marín, E., Floriano, B., Bernhardt, R., Ewen, K.M., Reyes-Ramírez, F. y col. (2011). ThnY is a ferredoxin reductase-like iron-sulfur flavoprotein that has evolved to function as a regulator of tetralin biodegradation gene expression. *The Journal of Biological Chemistry* 286:1709-1718.
- García, P.P., Bringhurst, R.M., Arango Pinedo, C. y Gage, D.J. (2010). Characterization of a two-component regulatory system that regulates succinate-mediated catabolite repression in *Sinorhizobium meliloti*. *Journal of Bacteriology* 192:5725-5735.
- García-Romero, I., Pérez-Pulido, A.J., González-Flores, Y.E., Reyes-Ramírez, F., Santero, E. y Floriano, B. (2016). Genomic analysis of the nitrate-respiring *Sphingopyxis granulii* (formerly *Sphingomonas macrogoltabida*) strain TFA. *BMC genomics* 17:93.
- Gardner, P.P., Daub, J., Tate, J., Moore, B.L., Osuch, I.H., Griffiths-Jones, S. y col. (2011). Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Research* 39:141.
- Garst, A.D., Edwards, A.L. y Batey, R.T. (2011). Riboswitches: structures and mechanisms. *Cold Spring Harbor Perspectives in Biology* 3:a003533.
- Gatheru Waigi, M., Sun, K. y Gao, Y. (2017). Sphingomonads in Microbe-Assisted Phytoremediation: Tackling Soil Pollution. *Trends in Biotechnology* 35:883-899.
- Ge Zhang and Feng Gao. (2017). Quantitative analysis of correlation between AT and GC biases among bacterial genomes. *PLoS One* 12:e0171408.
- Ghiglione, J., Richaume, A., Philippot, L. y Lensi, R. (2002). Relative involvement of nitrate and nitrite reduction in the competitiveness of *Pseudomonas fluorescens* in the rhizosphere of maize under non-limiting nitrate conditions. *FEMS microbiology ecology* 39:121-127.
- Ghosal, D., Ghosh, S., Dutta, T.K. y Ahn, Y. (2016). Current State of Knowledge in Microbial Degradation of Polycyclic Aromatic Hydrocarbons (PAHs): A Review. *Frontiers in Microbiology* 7:1369.
- Gibson, J. y S Harwood, C. (2002). Metabolic diversity in aromatic compound utilization by anaerobic microbes. *Annual Review of Microbiology* 56:345-369.
- Glaeser, S.P. y Kämpfer, P. (2014). The Family Sphingomonadaceae. En: Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E. y Thompson, F. (Eds.), *The Prokaryotes*, Springer Berlin Heidelberg 641-707.
- Goh, L., Purama, R.K. y Sudesh, K. (2014). Enhancement of stress tolerance in the polyhydroxyalkanoate producers without mobilization of the accumulated granules. *Applied Biochemistry and Biotechnology* 172:1585-1598.
- Görke, B. y Stülke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews. Microbiology* 6:613-624.
- Gouy, M., Guindon, S. y Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27:221-224.
- Grigoriev, A. (1998). Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Research* 26:2286-2290.



- Grissa, I., Vergnaud, G. y Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research* 35:52-57.
- Grosjean, H. y Björk, G.R. (2004). Enzymatic conversion of cytidine to lysidine in anticodon of bacterial isoleucyl-tRNA--an alternative way of RNA editing. *Trends in Biochemical Sciences* 29:165-168.
- Guillier, M. y Gottesman, S. (2008). The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. *Nucleic Acids Research* 36:6781-6794.
- Guimarães, L.C., Florczak-Wyspianska, J., de Jesús, L.B., Viana, M.V.C., Silva, A., Ramos, R.T.J.y col. (2015). Inside the Pan-genome - Methods and Software Overview. *Current Genomics* 16:245-252.
- Guo, M.S., Updegrave, T.B., Gogol, E.B., Shabalina, S.A., Gross, C.A. y Storz, G. (2014). MicL, a new  $\sigma$ E-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes & Development* 28:1620-1634.
- Gürtler, A., Kunz, N., Gomolka, M., Hornhardt, S., Friedl, A.A., McDonald, K.y col. (2013). Stain-Free technology as a normalization tool in Western blot analysis. *Analytical Biochemistry* 433:105-111.
- Guzmán, H., Van-Thuoc, D., Martín, J., Hatti-Kaul, R. y Quillaguamán, J. (2009). A process for the production of ectoine and poly(3-hydroxybutyrate) by *Halomonas boliviensis*. *Applied Microbiology and Biotechnology* 84:1069-1077.
- Haller, A., Altman, R.B., Soulière, M.F., Blanchard, S.C. y Micura, R. (2013). Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proceedings of the National Academy of Sciences of the United States of America* 110:4188-4193.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* 166:557-580.
- Hao, Y., Updegrave, T.B., Livingston, N.N. y Storz, G. (2016). Protection against deleterious nitrogen compounds: role of  $\sigma$ S-dependent small RNAs encoded adjacent to *sdhA*. *Nucleic Acids Research* 44:6935-6948.
- Hegedüs, B., Kós, P.B., Bálint, B., Maróti, G., Gan, H.M., Perei, K.y col. (2017). Complete genome sequence of *Novosphingobium resinovorum* SA1, a versatile xenobiotic-degrading bacterium capable of utilizing sulfanilic acid. *Journal of Biotechnology* 241:76-80.
- Henry, C.S., DeJongh, M., Best, A.A., Frybarger, P.M., Lindsay, B. y Stevens, R.L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology* 28:977-982.
- Hernández Silva, M.J. (2000), "Caracterización molecular de la ruta de degradación de tetralina en *Sphingomonas* sp. estirpe TFA", (Tesis, Universidad de Sevilla).
- Hernández, M.J., Andújar, E., Ríos, J.L., Kaschabek, S.R., Reineke, W. y Santero, E. (2000). Identification of a serine hydrolase which cleaves the alicyclic ring of tetralin. *Journal of Bacteriology* 182:5448-5453.

- Hernández, M.J., Floriano, B., Ríos, J.J. y Santero, E. (2002). Identification of a hydratase and a class II aldolase involved in biodegradation of the organic solvent tetralin. *Applied and Environmental Microbiology* 68:4841-4846.
- Hernández, M.J., Reineke, W. y Santero, E. (1999). Genetic analysis of biodegradation of tetralin by a *Sphingomonas* strain. *Applied and Environmental Microbiology* 65:1806-1810.
- Heroven, A.K., Sest, M., Pisano, F., Scheb-Wetzel, M., Steinmann, R., Böhme, K. y col. (2012). Crp induces switching of the CsrB and CsrC RNAs in *Yersinia pseudotuberculosis* and links nutritional status to virulence. *Frontiers in Cellular and Infection Microbiology* 2:158.
- Hirota-Mamoto, R., Nagai, R., Tachibana, S., Yasuda, M., Tani, A., Kimbara, K. y col. (2006). Cloning and expression of the gene for periplasmic poly(vinyl alcohol) dehydrogenase from *Sphingomonas* sp. strain 113P3, a novel-type quinoxalene alcohol dehydrogenase. *Microbiology* 152:1941-1949.
- Hoang, T.T., Karkhoff-Schweizer, R.R., Kutchma, A.J. y Schweizer, H.P. (1998). A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. *Gene* 212:77-86.
- Hsiao, W.W.L., Ung, K., Aeschliman, D., Bryan, J., Finlay, B.B. y Brinkman, F.S.L. (2005). Evidence of a large novel gene pool associated with prokaryotic genomic islands. *PLoS genetics* 1:e62.
- Hu, A., He, J., Chu, K.-. y Yu, C.-. (2011). Genome Sequence of the 17 -Estradiol-Utilizing Bacterium *Sphingomonas* Strain KC8. *Journal of Bacteriology* 193:4266-4267.
- Hu, X., Mamoto, R., Fujioka, Y., Tani, A., Kimbara, K. y Kawai, F. (2008). The *pva* operon is located on the megaplasmid of *Sphingopyxis* sp. strain 113P3 and is constitutively expressed, although expression is enhanced by PVA. *Applied Microbiology and Biotechnology* 78:685-693.
- Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H. y col. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19:524-531.
- Hugenholtz, P., Pitulle, C., Hershberger, K.L. y Pace, N.R. (1998). Novel division level bacterial diversity in a Yellowstone hot spring. *Journal of Bacteriology* 180:366-376.
- Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D. y col. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Research* 37:211.
- Hyatt, D., Chen, G., Locascio, P.F., Land, M.L., Larimer, F.W. y Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* 11:119.
- Imam, S., Yilmaz, S., Sohmen, U., Gorzalski, A.S., Reed, J.L., Noguera, D.R. y col. (2011). iRsp1095: a genome-scale reconstruction of the *Rhodobacter sphaeroides* metabolic network. *BMC systems biology* 5:116.
- Inoue, H., Nojima, H. y Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96:23-28.

- Ioannidis, P., Dunning Hotopp, J.C., Sapountzis, P., Siozios, S., Tsiamis, G., Bordenstein, S.R. y col. (2007). New criteria for selecting the origin of DNA replication in *Wolbachia* and closely related bacteria. *BMC genomics* 8:182.
- Jagodnik, J., Brosse, A., Le Lam, T.N., Chiaruttini, C. y Guillier, M. (2017). Mechanistic study of base-pairing small regulatory RNAs in bacteria. *Methods* 117:67-76.
- Jeffrey H. Miller. (1972). Experiments in molecular genetics. *Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.*
- Jühling, F., Mörl, M., Hartmann, R.K., Sprinzl, M., Stadler, P.F. y Pütz, J. (2009). tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Research* 37:159.
- Kanehisa, M. y Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28:27-30.
- Karp, P.D., Paley, S. y Romero, P. (2002a). The Pathway Tools software. *Bioinformatics* 18 Suppl 1:225.
- Karp, P.D., Riley, M., Paley, S.M. y Pellegrini-Toole, A. (2002b). The MetaCyc Database. *Nucleic Acids Research* 30:59-61.
- Kawai, F. (1999). Sphingomonads involved in the biodegradation of xenobiotic polymers. *Journal of Industrial Microbiology & Biotechnology* 23:400-407.
- Kawai, F. y Hu, X. (2009). Biochemistry of microbial polyvinyl alcohol degradation. *Applied Microbiology and Biotechnology* 84:227-237.
- Kawai, F., Kitajima, S., Oda, K., Higasa, T., Charoenpanich, J., Hu, X. y col. (2013). Polyvinyl alcohol and polyethylene glycol form polymer bodies in the periplasm of Sphingomonads that are able to assimilate them. *Archives of Microbiology* 195:131-140.
- Keseler, I.M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R. y col. (2017). The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Research* 45:550.
- Kim, B.C., Sohn, C.K., Lim, S.K., Lee, J.W. y Park, W. (2003). Degradation of polyvinyl alcohol by *Sphingomonas* sp. SA3 and its symbiote. *Journal of Industrial Microbiology & Biotechnology* 30:70-74.
- Kim, M.K., Im, W., Ohta, H., Lee, M. y Lee, S. (2005). *Sphingopyxis granuli* sp. nov., a beta-glucosidase-producing bacterium in the family Sphingomonadaceae in alpha-4 subclass of the Proteobacteria. *Journal of Microbiology* 43:152-157.
- Kinouchi, M. y Kurokawa, K. (2006). tRNAfinder: A software system to find all tRNA genes in the DNA sequence based on the cloverleaf secondary structure 7:116-124.
- Kisand, V. y Lettieri, T. (2013). Genome sequencing of bacteria: sequencing, de novo assembly and rapid analysis using open source tools. *BMC genomics* 14:211.
- Klappenbach, J.A., Dunbar, J.M. y Schmidt, T.M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Applied and Environmental Microbiology* 66:1328-1333.
- Klomklang, W., Tani, A., Kimbara, K., Mamoto, R., Ueda, T., Shimao, M. y col. (2005). Biochemical and molecular characterization of a periplasmic hydrolase for oxidized polyvinyl alcohol from *Sphingomonas* sp. strain 113P3. *Microbiology* 151:1255-1262.

- Köhler, C., Mandal, D., Gaston, K.W., Grosjean, H., Limbach, P.A. y Rajbhandary, U.L. (2014). Life without tRNA<sup>Ala</sup>-lysine synthetase: translation of the isoleucine codon AUA in *Bacillus subtilis* lacking the canonical tRNA<sup>Ala</sup>. *Nucleic Acids Research* 42:1904-1915.
- Kole, R. y Altman, S. (1979). Reconstitution of RNase P activity from inactive RNA and protein. *Proceedings of the National Academy of Sciences of the United States of America* 76:3795-3799.
- Konstantinidis, K.T. y Tiedje, J.M. (2005). Towards a genome-based taxonomy for prokaryotes. *Journal of Bacteriology* 187(18):6258-6264.
- Kornberg, A. y Baker, T.A. (1992). DNA replication. 2nd ed. New York: W. H. Freeman and Company. New York.
- Lagares, A., Roux, I. y Valverde, C. (2016). Phylogenetic distribution and evolutionary pattern of an  $\alpha$ -proteobacterial small RNA gene that controls polyhydroxybutyrate accumulation in *Sinorhizobium meliloti*. *Molecular Phylogenetics and Evolution* 99:182-193.
- Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H., Rognes, T. y Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research* 35:3100-3108.
- Land, M.L., Hyatt, D., Jun, S., Kora, G.H., Hauser, L.J., Lukjancenko, O. y col. (2014). Quality scores for 32,000 genomes. *Standards in Genomic Sciences* 9:20.
- Land, M., Hauser, L., Jun, S., Nookaew, I., Leuze, M.R., Ahn, T. y col. (2015). Insights from 20 years of bacterial genome sequencing. *Functional & Integrative Genomics* 15:141-161.
- Landt, S.G., Abeliuk, E., McGrath, P.T., Lesley, J.A., McAdams, H.H. y Shapiro, L. (2008). Small non-coding RNAs in *Caulobacter crescentus*. *Molecular Microbiology* 68:600-614.
- Laslett, D. y Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research* 32:11-16.
- Lauro, F.M., McDougald, D., Thomas, T., Williams, T.J., Egan, S., Rice, S. y col. (2009). The genomic basis of trophic strategy in marine bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 106:15527-15533.
- Laursen, B.S., Sørensen, H.P., Mortensen, K.K. y Sperling-Petersen, H.U. (2005). Initiation of protein synthesis in bacteria. *Microbiology and molecular biology reviews: MMBR* 69:101-123.
- Ledesma-García, L., Reyes-Ramírez, F. y Santero, E. (2013). The ferredoxin ThnA3 negatively regulates tetralin biodegradation gene expression via ThnY, a ferredoxin reductase that functions as a regulator of the catabolic pathway. *PloS One* 8:e73910.
- Ledesma-García, L., Sánchez-Azqueta, A., Medina, M., Reyes-Ramírez, F. y Santero, E. (2016). Redox proteins of hydroxylating bacterial dioxygenases establish a regulatory cascade that prevents gratuitous induction of tetralin biodegradation genes. *Scientific Reports* 6:23848.
- Lee, T.J., Paulsen, I. y Karp, P. (2008). Annotation-based inference of transporter function. *Bioinformatics* 24:259-267.

- Lewis, N.E., Nagarajan, H. y Palsson, B.O. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature reviews Microbiology* 10:291-305.
- Li, H. y Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. y col. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079.
- Li, L., Huang, D., Cheung, M.K., Nong, W., Huang, Q. y Kwan, H.S. (2013). BSRD: a repository for bacterial small regulatory RNA. *Nucleic Acids Research* 41:233.
- Liao, Y., Tsai, M., Chen, F. y Hsiung, C.A. (2012). GEMSiRV: a software platform for GENome-scale metabolic model simulation, reconstruction and visualization. *Bioinformatics* 28:1752-1758.
- Liu, M., Durfee, T., Cabrera, J.E., Zhao, K., Jin, D.J. y Blattner, F.R. (2005). Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*. *The Journal of Biological Chemistry* 280:15921-15927.
- Lobry, J.R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular Biology and Evolution* 13:660-665.
- Loewenstein, Y., Raimondo, D., Redfern, O.C., Watson, J., Frishman, D., Linial, M. y col. (2009). Protein function annotation by homology-based inference. *Genome Biology* 10:207.
- Loman, N.J. y Pallen, M.J. (2015). Twenty years of bacterial genome sequencing. *Nature Reviews. Microbiology* 13:787-794.
- López-Sánchez, A., Floriano, B., Andújar, E., Hernáez, M.J. y Santero, E. (2010). Tetralin-induced and ThnR-regulated aldehyde dehydrogenase and beta-oxidation genes in *Sphingomonas macrogolita* strain TFA. *Applied and Environmental Microbiology* 76:110-118.
- López-Sánchez, A., Rivas-Marín, E., Martínez-Pérez, O., Floriano, B. y Santero, E. (2009). Co-ordinated regulation of two divergent promoters through higher-order complex formation by the LysR-type regulator ThnR. *Molecular Microbiology* 73:1086-1100.
- Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F. y col. (2011). ViennaRNA Package 2.0. *Algorithms for molecular biology* 6:26.
- Lowe, T.M. y Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25:955-964.
- Lutz, R. y Bujard, H. (1997). Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research* 25:1203-1210.
- Macé, K. y Gillet, R. (2016). Origins of tmRNA: the missing link in the birth of protein synthesis? *Nucleic Acids Research* 44:8041-8051.
- Macián, F., Pérez-Roger, I. y Armengod, M.E. (1994). An improved vector system for constructing transcriptional *lacZ* fusions: analysis of regulation of the *dnaA*, *dnaN*, *recF* and *gyrB* genes of *Escherichia coli*. *Gene* 145:17-24.

## Bibliografía

- Mackiewicz, P., Zakrzewska-Czerwinska, J., Zawilak, A., Dudek, M.R. y Cebrat, S. (2004). Where does bacterial replication start? Rules for predicting the oriC region. *Nucleic Acids Research* 32:3781-3791.
- Macur, R.E., Wheeler, J.T., McDermott, T.R. y Inskeep, W.P. (2001). Microbial populations associated with the reduction and enhanced mobilization of arsenic in mine tailings. *Environmental Science & Technology* 35:3676-3682.
- Madhugiri, R., Pessi, G., Voss, B., Hahn, J., Sharma, C.M., Reinhardt, R. y col. (2012). Small RNAs of the *Bradyrhizobium/Rhodopseudomonas* lineage and their analysis. *RNA biology* 9:47-58.
- Majdalani, N., Vanderpool, C.K. y Gottesman, S. (2005). Bacterial small RNA regulators. *Critical Reviews in Biochemistry and Molecular Biology* 40:93-113.
- Malde, K. (2011). Flower: extracting information from pyrosequencing data. *Bioinformatics* 27:1041-1042.
- Mamoto, R., Hu, X., Chiue, H., Fujioka, Y. y Kawai, F. (2008). Cloning and expression of soluble cytochrome c and its role in polyvinyl alcohol degradation by polyvinyl alcohol-utilizing *Sphingopyxis* sp. strain 113P3. *Journal of Bioscience and Bioengineering* 105:147-151.
- Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L. y col. (2004). A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 306:275-279.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bembien, L.A. y col. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376-380.
- Martens, M., Dawyndt, P., Coopman, R., Gillis, M., De Vos, P. y Willems, A. (2008). Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). *International Journal of Systematic and Evolutionary Microbiology* 58:200-214.
- Martín-Cabello, G., Moreno-Ruiz, E., Morales, V., Floriano, B. y Santero, E. (2011). Involvement of poly(3-hydroxybutyrate) synthesis in catabolite repression of tetralin biodegradation genes in *Sphingomonas macrogolita* strain TFA. *Environmental Microbiology Reports* 3:627-631.
- Martínez-Pérez, O. (2008), "Regulación de los genes de degradación de tetralina" (Tesis, Universidad de Sevilla).
- Martínez-Pérez, O., Moreno-Ruiz, E., Floriano, B. y Santero, E. (2004). Regulation of tetralin biodegradation and identification of genes essential for expression of the thn operons. *Journal of Bacteriology* 186:6101-6109.
- Martínez-Pérez, O., López-Sánchez, A., Reyes-Ramírez, F., Floriano, B. y Santero, E. (2007). Integrated response to inducers by communication between a catabolic pathway and its regulatory system. *Journal of Bacteriology* 189:3768-3775.
- Massé, E. y Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 99:4620-4625.

- Matte, A., Ekiel, I., Jia, Z., Gehring, K. y Cygler, M. (2009). A medium-throughput structural proteomics approach applied to the genome of *E. coli*. En: Lee S.Y. (eds) *Systems Biology and Biotechnology of Escherichia coli*, Springer, Dordrecht. 59-86.
- Mattinen, L., Somervuo, P., Nykyri, J., Nissinen, R., Kouvonen, P., Corthals, G. y col. (2008). Microarray profiling of host-extract-induced genes and characterization of the type VI secretion cluster in the potato pathogen *Pectobacterium atrosepticum*. *Microbiology* 154:2387-2396.
- McCarthy, J.E. y Brimacombe, R. (1994). Prokaryotic translation: the interactive pathway leading to initiation. *Trends in genetics* 10:402-407.
- McCloskey, D., Palsson, B.Ø and Feist, A.M. (2013). Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Molecular Systems Biology* 9:661.
- Médigue, C. y Moszer, I. (2007). Annotation, comparison and databases for hundreds of bacterial genomes. *Research in Microbiology* 158:724-736.
- Miller, J.R., Koren, S. y Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics* 95:315-327.
- Mira, A., Martín-Cuadrado, A.B., D'Auria, G. y Rodríguez-Valera, F. (2010). The bacterial pan-genome: a new paradigm in microbiology. *International Microbiology: The Official Journal of the Spanish Society for Microbiology* 13:45-57.
- Miyakoshi, M., Chao, Y. y Vogel, J. (2015). Regulatory small RNAs from the 3' regions of bacterial mRNAs. *Current Opinion in Microbiology* 24:132-139.
- Moreno, R., Fonseca, P. y Rojo, F. (2012). Two small RNAs, CrcY and CrcZ, act in concert to sequester the Crc global regulator in *Pseudomonas putida*, modulating catabolite repression. *Molecular Microbiology* 83:24-40.
- Moreno, R., Hernández-Arranz, S., La Rosa, R., Yuste, L., Madhushani, A., Shingler, V. y col. (2015). The Crc and Hfq proteins of *Pseudomonas putida* cooperate in catabolite repression and formation of ribonucleic acid complexes with specific target motifs. *Environmental Microbiology* 17:105-118.
- Moreno-Ruiz, E., Hernáez, M.J., Martínez-Pérez, O. y Santero, E. (2003). Identification and functional characterization of *Sphingomonas macroglutabida* strain TFA genes involved in the first two steps of the tetralin catabolic pathway. *Journal of Bacteriology* 185:2026-2030.
- Morita, T., Nishino, R. y Aiba, H. (2017). Role of the terminator hairpin in the biogenesis of functional Hfq-binding sRNAs. *RNA* 23:1419-1431.
- Moritz, K.D., Amendt, B., Witt, Elisabeth M H J and Galinski, E.A. (2015). The hydroxyectoine gene cluster of the non-halophilic acidophile *Acidiphilium cryptum*. *Extremophiles: Life Under Extreme Conditions* 19:87-99.
- Motamedian, E., Saeidi, M. y Shojaosadati, S.A. (2016). Reconstruction of a charge balanced genome-scale metabolic model to study the energy-uncoupled growth of *Zymomonas mobilis* ZM1. *Molecular bioSystems* 12:1241-1249.
- Muñoz-Mérida, A., Viguera, E., Claros, M.G., Trelles, O. y Pérez-Pulido, A.J. (2014). Sma3s: a three-step modular annotator for large sequence datasets. *DNA research: an international journal for rapid publication of reports on genes and genomes* 21:341-353.

## Bibliografía

- Murphy, F.V. y Ramakrishnan, V. (2004). Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nature Structural & Molecular Biology* 11:1251-1252.
- Muzzi, A., Masignani, V. y Rappuoli, R. (2007). The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. *Drug Discovery Today* 12:429-439.
- Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J. y col. (2000). A whole-genome assembly of *Drosophila*. *Science* 287:2196-2204.
- Nagata, S., Wang, Y., Oshima, A., Zhang, L., Miyake, H., Sasaki, H. y col. (2008). Efficient cyclic system to yield ectoine using *Brevibacterium* sp. JCM 6894 subjected to osmotic downshock. *Biotechnology and Bioengineering* 99:941-948.
- Nagata, Y., Kamakura, M., Endo, R., Miyazaki, R., Ohtsubo, Y. y Tsuda, M. (2006). Distribution of gamma-hexachlorocyclohexane-degrading genes on three replicons in *Sphingobium japonicum* UT26. *FEMS microbiology letters* 256:112-118.
- Nawrocki, E.P. (2014). Annotating functional RNAs in genomes using Infernal. *Methods in Molecular Biology* 1097:163-197.
- Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R. y col. (2015). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research* 43:130.
- Nawrocki, E.P. y Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29:2933-2935.
- Nawrocki, E.P., Kolbe, D.L. y Eddy, S.R. (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25:1335-1337.
- Nogales, J. (2014). A practical protocol for genome-scale metabolic reconstructions. En: McGenity T., Timmis K., y Nogales B. (eds). *Hydrocarbon and Lipid Microbiology Protocols*, Springer, Berlin, Heidelberg. 197-221.
- Nogales, J. y Agudo, L. (2015). A practical protocol for integration of transcriptomics data into genome-scale metabolic reconstructions. En: McGenity T., Timmis K., y Nogales B. (eds). *Hydrocarbon and Lipid Microbiology Protocols*, Springer, Berlin, Heidelberg. 135-152.
- Nogales, J., Gudmundsson, S., Duque, E., Ramos, J.L. y Palsson, B.O. (2017). Expanding The Computable Reactome in *Pseudomonas putida* Reveals Metabolic Cycles Providing Robustness. *bioRxiv*.
- Nogales, J., Palsson, B.Ø and Thiele, I. (2008). A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. *BMC systems biology* 2:79.
- Oberto, J. (2013). SyntTax: a web server linking synteny to prokaryotic taxonomy. *BMC bioinformatics* 14:4.
- Obruca, S., Sedlacek, P., Mravec, F., Samek, O. y Marova, I. (2016). Evaluation of 3-hydroxybutyrate as an enzyme-protective agent against heating and oxidative damage and its potential role in stress response of poly(3-hydroxybutyrate) accumulating cells. *Applied Microbiology and Biotechnology* 100:1365-1376.
- Omotajo, D., Tate, T., Cho, H. y Choudhary, M. (2015). Distribution and diversity of ribosome binding sites in prokaryotic genomes. *BMC genomics* 16:604.



- Opdyke, J.A., Kang, J. y Storz, G. (2004). GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *Journal of Bacteriology* 186:6698-6705.
- Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M. y col. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism--2011. *Molecular Systems Biology* 7:535.
- Ozaki, S. y Katayama, T. (2012). Highly organized DnaA-oriC complexes recruit the single-stranded DNA for replication initiation. *Nucleic Acids Research* 40:1648-1665.
- Palsson, B.Ø, Orth, J.D. y Thiele, I. (2010). What is flux balance analysis? *Nature Biotechnology* 28:245-248.
- Pannuri, A., Vakulskas, C.A., Zere, T., McGibbon, L.C., Edwards, A.N., Georgellis, D. y col. (2016). Circuitry Linking the Catabolite Repression and Csr Global Regulatory Systems of *Escherichia coli*. *Journal of Bacteriology* 198:3000-3015.
- Papenfort, K., Bouvier, M., Mika, F., Sharma, C.M. y Vogel, J. (2010). Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proceedings of the National Academy of Sciences of the United States of America* 107:20435-20440.
- Papenfort, K., Podkaminski, D., Hinton, J.C.D. y Vogel, J. (2012). The ancestral SgrS RNA discriminates horizontally acquired *Salmonella* mRNAs through a single G-U wobble pair. *Proceedings of the National Academy of Sciences of the United States of America* 109:757.
- Pareek, C.S., Smoczynski, R. y Tretyn, A. (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics* 52:413-435.
- Pastor, J.M., Salvador, M., Argandoña, M., Bernal, V., Reina-Bueno, M., Csonka, L.N. y col. (2010). Ectoines in cell stress protection: uses and biotechnological production. *Biotechnology Advances* 28:782-801.
- Pavez, P., Castillo, J.L., González, C. y Martínez, M. (2009). Poly-beta-hydroxyalkanoate exert a protective effect against carbon starvation and frozen conditions in *Sphingopyxis chilensis*. *Current Microbiology* 59:636-640.
- Pédrelacq, J., Cabantous, S., Tran, T., Terwilliger, T.C. y Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* 24:79-88.
- Petrov, A.S., Bernier, C.R., Gulen, B., Waterbury, C.C., HersHKovits, E., Hsiao, C. y col. (2014). Secondary structures of rRNAs from all three domains of life. *PLoS One* 9:e88222.
- Peyraud, R., Schneider, K., Kiefer, P., Massou, S., Vorholt, J.A. y Portais, J. (2011). Genome-scale reconstruction and system level investigation of the metabolic network of *Methylobacterium extorquens* AM1. *BMC systems biology* 5:189.
- Pieper, D.H. y Reineke, W. (2000). Engineering bacteria for bioremediation. *Current Opinion in Biotechnology* 11:262-270.
- Pinedo, C.A., Bringham, R.M. y Gage, D.J. (2008). *Sinorhizobium meliloti* mutants lacking phosphotransferase system enzyme HPr or EIIA are altered in diverse processes, including carbon metabolism, cobalt requirements, and succinoglycan production. *Journal of Bacteriology* 190:2947-2956.

## Bibliografía

- Pinedo, C.A. y Gage, D.J. (2009). HPrK regulates succinate-mediated catabolite repression in the gram-negative symbiont *Sinorhizobium meliloti*. *Journal of Bacteriology* 191:298-309.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. y col. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41:590.
- Raman, M.C.C., Johnson, K.A., Clarke, D.J., Naismith, J.H. y Campopiano, D.J. (2010). The serine palmitoyltransferase from *Sphingomonas wittichii* RW1: An interesting link to an unusual acyl carrier protein. *Biopolymers* 93:811-822.
- Reed, J.L., Vo, T.D., Schilling, C.H. y Palsson, B.O. (2003). An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biology* 4:R54.
- Resendis-Antonio, O., Hernández, M., Mora, Y. y Encarnación, S. (2012). Functional modules, structural topology, and optimal activity in metabolic networks. *PLoS computational biology* 8:e1002720.
- Richter, M. y Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* 106:19126-19131.
- Richter, M., Rosselló-Móra, R., Oliver Glöckner, F. y Peplies, J. (2016). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32:929-931.
- Rivas, E. y Eddy, S.R. (2001). Noncoding RNA gene detection using comparative sequence analysis. *BMC bioinformatics* 2:8.
- Roberts, M.F. (2005). Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline Systems* 1:5.
- Rocha, E.P.C. (2004). Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Research* 14:2279-2286.
- Rojo, F. (2010). Carbon catabolite repression in *Pseudomonas*: optimizing metabolic versatility and interactions with the environment. *FEMS microbiology reviews* 34:658-684.
- Romine, M.F., Stillwell, L.C., Wong, K.K., Thurston, S.J., Sisk, E.C., Sensen, C. y col. (1999). Complete sequence of a 184-kilobase catabolic plasmid from *Sphingomonas aromaticivorans* F199. *Journal of Bacteriology* 181:1585-1602.
- Rothberg, J.M. y Leamon, J.H. (2008). The development and impact of 454 sequencing. *Nature Biotechnology* 26:1117-1124.
- Sadykov, M.R., Ahn, J., Widhelm, T.J., Eckrich, V.M., Endres, J.L., Driks, A. y col. (2017). Poly(3-hydroxybutyrate) fuels the tricarboxylic acid cycle and de novo lipid biosynthesis during *Bacillus anthracis* sporulation. *Molecular Microbiology* 104:793-803.
- Saitou, N. y Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
- Sakazawa, C., Shimao, M., Taniguchi, Y. y Kato, N. (1981). Symbiotic utilization of polyvinyl alcohol by mixed cultures. *Applied and Environmental Microbiology* 41:261-267.

- Salzberg, S.L., Delcher, A.L., Kasif, S. y White, O. (1998). Microbial gene identification using interpolated Markov models. *Nucleic Acids Research* 26:544-548.
- Sambrook, J., Fritsch, E.F. y Maniatis, T. (1989). Molecular cloning: a laboratory manual. New York : Cold Spring Harbor Laboratory.
- Sanger, F. y Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* 94:441-448.
- Sanger, F., Nicklen, S. y Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74:5463-5467.
- Santos, P.M., Mignogna, G., Heipieper, H.J. and Zennaro, E. (2002). Occurrence and properties of glutathione S-transferases in phenol-degrading *Pseudomonas* strains. *Research in Microbiology* 153:89-98.
- Schellenberger, J., Park, J.O., Conrad, T.M. y Palsson, B.Ø. (2010). BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC bioinformatics* 11:213.
- Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M.y col. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols* 6:1290-1307.
- Schomburg, I., Chang, A. y Schomburg, D. (2002). BRENDA, enzyme data and metabolic information. *Nucleic Acids Research* 30:47-49.
- Schreiber, A.F. y Winkler, U.K. (1983). Transformation of tetralin by whole cells of *Pseudomonas stutzeri* AS39. *European journal of applied microbiology and biotechnology* 18:6-10.
- Schut, F., Jansen, M., Gomes, T.M., Gottschal, J.C., Harder, W. y Prins, R.A. (1995). Substrate uptake and utilization by a marine ultramicrobacterium. *Microbiology* 141 ( Pt 2):351-361.
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068-2069.
- Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A.y col. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464:250-255.
- Shimao, M. (2001). Biodegradation of plastics. *Current Opinion in Biotechnology* 12:242-247.
- Sikkema, J. y de Bont, J.A. (1993). Metabolism of tetralin (1,2,3,4-tetrahydronaphthalene) in *Corynebacterium* sp. strain C125. *Applied and Environmental Microbiology* 59:567-572.
- Sikkema, J., Poolman, B., Konings, W.N. y de Bont, J.A. (1992). Effects of the membrane action of tetralin on the functional and structural properties of artificial and bacterial membranes. *Journal of Bacteriology* 174:2986-2992.
- Sikkema, J. y Bont, Jan A M de. (1991). Isolation and initial characterization of bacteria growing on tetralin. *Biodegradation* 2:15-23.
- Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T.y col. (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS genetics* 4:e1000163.

## Bibliografía

- Sittka, A., Pfeiffer, V., Tedin, K. y Vogel, J. (2007). The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Molecular Microbiology* 63:193-217.
- Sivashankari, S. y Shanmughavel, P. (2007). Comparative genomics - a perspective. *Bioinformatics* 1:376-378.
- Snipen, L., Almøy, T. y Ussery, D.W. (2009). Microbial comparative pan-genomics using binomial mixture models. *BMC genomics* 10:385.
- Solovyev, V. y Salamov, A. (2011). Automatic annotation of microbial genomes and metagenomic sequences. En: *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*, (Ed. R.W. Li), Nova Science Publishers. 61-78.
- Sonnleitner, E. y Bläsi, U. (2014). Regulation of Hfq by the RNA CrcZ in *Pseudomonas aeruginosa* carbon catabolite repression. *PLoS genetics* 10:e1004440.
- Sridhar, J. y Gunasekaran, P. (2013). Computational small RNA prediction in bacteria. *Bioinformatics and Biology Insights* 7:83-95.
- Staskawicz, B., Dahlbeck, D., Keen, N. y Napoli, C. (1987). Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv. *glycinea*. *Journal of Bacteriology* 169:5789-5794.
- Stephen, D., Jones, C. y Schofield, J.P. (1990). A rapid method for isolating high quality plasmid DNA suitable for DNA sequencing. *Nucleic Acids Research* 18:7463-7464.
- Stolz, A. (2009). Molecular characteristics of xenobiotic-degrading sphingomonads. *Applied Microbiology and Biotechnology* 81:793-811.
- Stolz, A. (2014). Degradative plasmids from sphingomonads. *FEMS microbiology letters* 350:9-19.
- Storz, G., Vogel, J. y Wassarman, K.M. (2011). Regulation by small RNAs in bacteria: expanding frontiers. *Molecular Cell* 43:880-891.
- Stothard, P. y Wishart, D.S. (2006). Automated bacterial genome analysis and annotation. *Current Opinion in Microbiology* 9:505-510.
- Suzuki, T., Ichihara, Y., Yamada, M. y Tonomura, K. (1973). Some Characteristics of *Pseudomonas* O-3 which Utilizes Polyvinyl Alcohol. *Agricultural and Biological Chemistry* 37:747-756.
- Svensson, S.L. y Sharma, C.M. (2016). Small RNAs in Bacterial Virulence and Communication. *Microbiology Spectrum* 4:VMBF-0028-2015.
- Taghavi, S., van der Lelie, D., Hoffman, A., Zhang, Y., Walla, M.D., Vangronsveld, J. y col. (2010). Genome sequence of the plant growth promoting endophytic bacterium *Enterobacter* sp. 638. *PLoS genetics* 6:e1000943.
- Takeuchi, M., Hamana, K. y Hiraishi, A. (2001). Proposal of the genus *Sphingomonas sensu stricto* and three new genera, *Sphingobium*, *Novosphingobium* and *Sphingopyxis*, on the basis of phylogenetic and chemotaxonomic analyses. *International Journal of Systematic and Evolutionary Microbiology* 51:1405-1417.
- Tettelin, H., Maignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L. y col. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications

- for the microbial "pan-genome". *Proceedings of the National Academy of Sciences of the United States of America* 102:13950-13955.
- The UniProt Consortium. (2007). The Universal Protein Resource (UniProt). *Nucleic Acids Research* 35:193.
- The UniProt Consortium. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Research* 45:D169.
- Thiele, I. y Palsson, B.Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols* 5:93-121.
- Thompson, C.C., Chimetto, L., Edwards, R.A., Swings, J., Stackebrandt, E. y Thompson, F.L. (2013). Microbial genomic taxonomy. *BMC genomics* 14:913.
- Thorvaldsdóttir, H., Robinson, J.T. y Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* 14:178-192.
- Tucker, B.J. y Breaker, R.R. (2005). Riboswitches as versatile gene control elements. *Current Opinion in Structural Biology* 15:342-348.
- Updegrave, T.B., Zhang, A. y Storz, G. (2016). Hfq: the flexible RNA matchmaker. *Current Opinion in Microbiology* 30:133-138.
- Van Domselaar, G.H., Stothard, P., Shrivastava, S., Cruz, J.A., Guo, A., Dong, X. y col. (2005). BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Research* 33:455.
- Vanderpool, C.K., Balasubramanian, D. y Lloyd, C.R. (2011). Dual-function RNA regulators in bacteria. *Biochimie* 93:1943-1949.
- Vesth, T., Lagesen, K., Acar, Ö and Ussery, D. (2013). CMG-biotools, a free workbench for basic comparative microbial genomics. *PloS One* 8:e60120.
- Vitreschak, A.G., Rodionov, D.A., Mironov, A.A. y Gelfand, M.S. (2003). Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *RNA* 9:1084-1097.
- Vogel, J. y Papenfort, K. (2006). Small non-coding RNAs and the bacterial outer membrane. *Current Opinion in Microbiology* 9:605-611.
- Voorhees, R.M., Mandal, D., Neubauer, C., Köhrer, C., RajBhandary, U.L. y Ramakrishnan, V. (2013). The structural basis for specific decoding of AUA by isoleucine tRNA on the ribosome. *Nature Structural & Molecular Biology* 20:641-643.
- Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W.F. y col. (2006). Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC bioinformatics* 7:142.
- Wade, J.T. y Grainger, D.C. (2014). Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nature Reviews. Microbiology* 12:647-653.
- Wagner, E.G.H. y Romby, P. (2015). Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Advances in Genetics* 90:133-208.

## Bibliografía

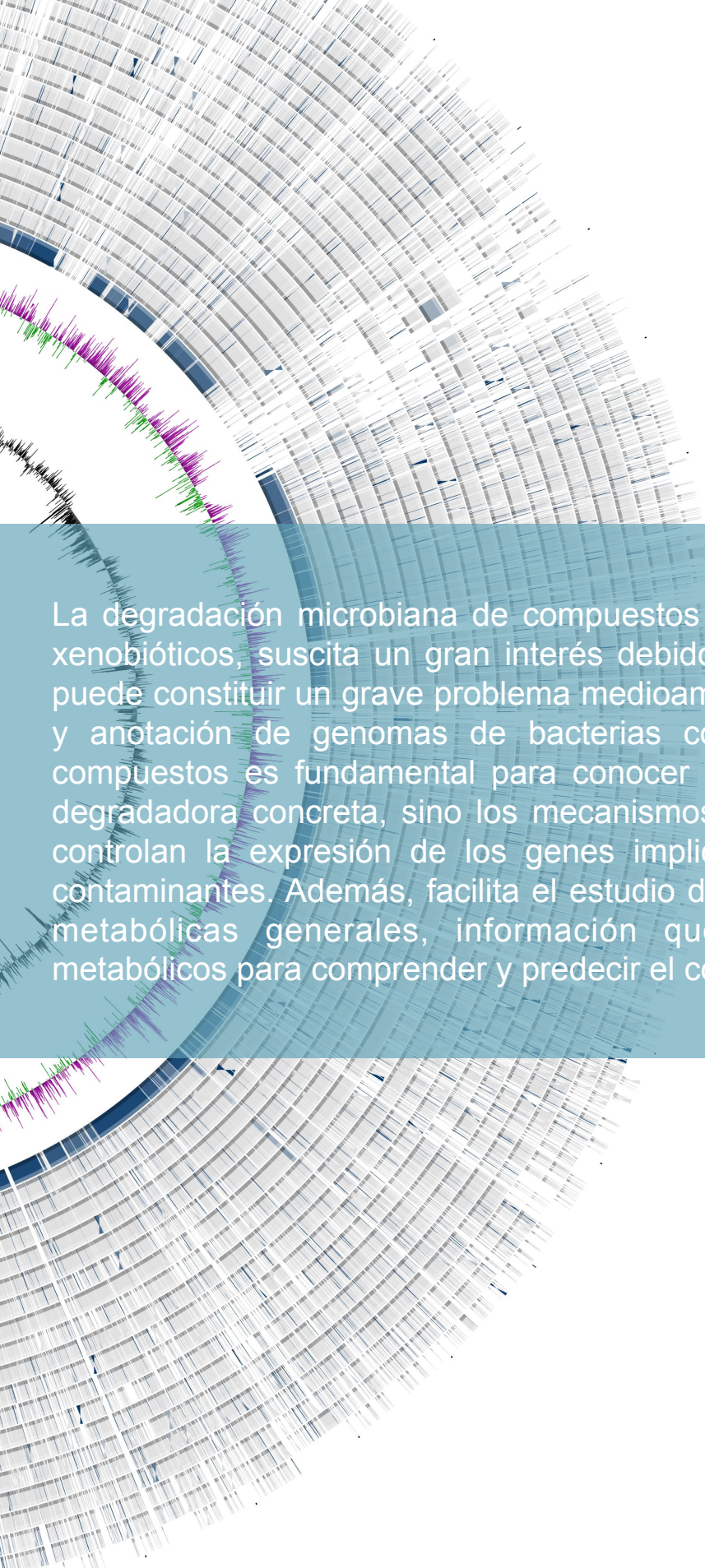
- Wang, Q., Yu, H., Xia, Y., Kang, Z. y Qi, Q. (2009). Complete PHB mobilization in *Escherichia coli* enhances the stress tolerance: a potential biotechnological application. *Microbial Cell Factories* 8:47.
- Wassarman, K.M. y Storz, G. (2000). 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* 101:613-623.
- Waters, L.S. y Storz, G. (2009). Regulatory RNAs in bacteria. *Cell* 136:615-628.
- Weinberg, Z., Wang, J.X., Bogue, J., Yang, J., Corbino, K., Moy, R.H. y col. (2010). Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biology* 11:R31.
- Williams, T.J., Ertan, H., Ting, L. y Cavicchioli, R. (2009). Carbon and nitrogen substrate utilization in the marine bacterium *Sphingopyxis alaskensis* strain RB2256. *The ISME journal* 3:1036-1052.
- Winsor, G.L., Lam, D.K.W., Fleming, L., Lo, R., Whiteside, M.D., Yu, N.Y. y col. (2011). *Pseudomonas* Genome Database: improved comparative analysis and population genomics capability for *Pseudomonas* genomes. *Nucleic Acids Research* 39:596.
- Wittich, R.M., Wilkes, H., Sinnwell, V., Francke, W. y Fortnagel, P. (1992). Metabolism of dibenzo-p-dioxin by *Sphingomonas* sp. strain RW1. *Applied and Environmental Microbiology* 58:1005-1010.
- Wolf, J., Gerber, A.P. y Keller, W. (2002). *tadA*, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *The EMBO journal* 21:3841-3851.
- Worning, P., Jensen, L.J., Hallin, P.F., Staerfeldt, H. y Ussery, D.W. (2006). Origin of replication in circular prokaryotic chromosomes. *Environmental Microbiology* 8:353-361.
- Wu, D., He, J., Gong, Y., Chen, D., Zhu, X., Qiu, N. y col. (2011a). Proteomic analysis reveals the strategies of *Bacillus thuringiensis* YBT-1520 for survival under long-term heat stress. *Proteomics* 11:2580-2591.
- Wu, S., Zhu, Z., Fu, L., Niu, B. y Li, W. (2011b). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC genomics* 12:444.
- Wu, X., Wang, X. y Lu, W. (2014). Genome-scale reconstruction of a metabolic network for *Gluconobacter oxydans* 621H. *Bio Systems* 117:10-14.
- Xia, X. (2012). DNA replication and strand asymmetry in prokaryotic and mitochondrial genomes. *Current Genomics* 13:16-27.
- Xu, D. y Xu, Y. (2004). Protein databases on the internet. *Current Protocols in Protein Science* Chapter 2, Unit 2.6.
- Yabuuchi, E. y Kosako, Y. (2015). *Sphingomonadales* ord. nov. En: *Bergey's Manual of Systematics of Archaea and Bacteria*, John Wiley & Sons, Inc., en asociación con Bergey's Manual Trust. 1-6.
- Yamatsu, A., Matsumi, R., Atomi, H. y Imanaka, T. (2006). Isolation and characterization of a novel poly(vinyl alcohol)-degrading bacterium, *Sphingopyxis* sp. PVA3. *Applied Microbiology and Biotechnology* 72:804-811.

- Yang, Y., Ko, T., Liu, L., Li, J., Huang, C., Chan, H.y col. (2014). Structural insights into enzymatic degradation of oxidized polyvinyl alcohol. *Chembiochem: A European Journal of Chemical Biology* 15:1882-1886.
- Yanisch-Perron, C., Vieira, J. y Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33:103-119.
- Yokobori, S., Kitamura, A., Grosjean, H. y Bessho, Y. (2013). Life without tRNA<sup>Arg</sup>-adenosine deaminase TadA: evolutionary consequences of decoding the four CGN codons as arginine in *Mycoplasmas* and other *Mollicutes*. *Nucleic Acids Research* 41:6531-6543.
- Yoon, S., Ha, S., Kwon, S., Lim, J., Kim, Y., Seo, H.y col. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International Journal of Systematic and Evolutionary Microbiology* 67:1613-1617.
- Zhang, L., Gao, J., Kim, S., Zhang, C., Jiang, J., Ma, X.y col. (2016). *Novosphingobium oryzae* sp. nov., a potential plant-promoting endophytic bacterium isolated from rice roots. *International Journal of Systematic and Evolutionary Microbiology* 66:302-307.
- Zhang, R. y Zhang, C. (2014). A Brief Review: The Z-curve Theory and its Application in Genome Analysis. *Current Genomics* 15:78-94.
- Zhao, H., Li, M., Fang, K., Chen, W. y Wang, J. (2012). In silico insights into the symbiotic nitrogen fixation in *Sinorhizobium meliloti* via metabolic reconstruction. *PloS One* 7:e31287.
- Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J. y Wishart, D.S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Research* 39:347.
- Zou, W., Liu, L., Zhang, J., Yang, H., Zhou, M., Hua, Q. y col. (2012). Reconstruction and analysis of a genome-scale metabolic model of the vitamin C producing industrial strain *Ketogulonicigenium vulgare* WSH-001. *Journal of Biotechnology* 161:42-48.
- Zwieb, C., van Nues, R.W., Rosenblad, M.A., Brown, J.D. y Samuelsson, T. (2005). A nomenclature for all signal recognition particle RNAs. *RNA* 11:7-13.









La degradación microbiana de compuestos recalcitrantes, tanto naturales como xenobióticos, suscita un gran interés debido a que su persistencia en el medio puede constituir un grave problema medioambiental y de salud. La secuenciación y anotación de genomas de bacterias con la habilidad de degradar estos compuestos es fundamental para conocer en más detalle no ya su capacidad degradadora concreta, sino los mecanismos globales de regulación génica que controlan la expresión de los genes implicados en el catabolismo de dichos contaminantes. Además, facilita el estudio de su fisiología y de sus capacidades metabólicas generales, información que puede integrarse en modelos metabólicos para comprender y predecir el comportamiento de estas bacterias.

**UNIVERSIDAD PABLO DE OLAVIDE-CSIC**

**Departamento de Biología Molecular e Ingeniería Bioquímica**



UNIVERSIDAD  
**PABLO DE OLAVIDE**  
SEVILLA



CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

**CSIC**



JUNTA DE ANDALUCÍA